

EPIGENETIC CHROMOSOME INTERACTIONS

Field of the Invention

The invention relates to detecting chromosome interactions and rheumatoid arthritis.

- 5 More particularly, the invention relates to a method of determining responsiveness to a specific therapy for rheumatoid arthritis in a subject; a companion diagnostic method; a therapeutic agent for use in the treatment and/or prophylaxis of rheumatoid arthritis in an individual (in particular in a human individual); a method of screening for (identifying) an agent, in particular a therapeutic agent, which is capable of changing responsiveness (in particular of an individual e.g. human individual) to a
- 10 therapy for rheumatoid arthritis; a method of determining the effect of a drug (e.g. therapeutic agent) comprising detecting the change in epigenetic chromosome interactions caused by the drug; and/or a library of nucleic acid and/or a nucleic acid.

Background of the Invention

15 Healthcare costs are spiralling and so there is a need to treat people more effectively using existing drugs. Some patients are non-responsive to particular pharmaceutical treatments. One example is the treatment of rheumatoid arthritis by methotrexate (MTX).

Rheumatoid arthritis (RA) is a chronic autoimmune disease affecting up to 1% of the global population.

- 20 Pathogenesis is multifactorial and characterized by primarily immune host gene loci interacting with environmental factors, particularly smoking and other pulmonary stimuli^{1,2,3}. The exposure of a genetically susceptible individual to such environmental factors suggests an epigenetic context for disease onset and progression. Recent studies of chromatin markers (e.g. methylation status of the genome) provide the first evidence of epigenetic differences associated with RA^{4,5,6,7}. However, to
- 25 date neither genetic associations, nor epigenetic changes, have provided a validated predictive marker for response to a given therapy. Moreover, clinical presentation only weakly predicts the efficacy and toxicity of known disease modifying anti-rheumatic drugs (DMARDs) such as methotrexate (MTX).

- MTX⁸, the commonest first-choice medication recommended by EULAR and ACR management
- 30 guidelines, delivers clinically meaningful response rates ranging from 50% to 65% after 6 months of treatment¹¹. Such responses, and especially the rather smaller proportion that exhibits high hurdle responses, cannot currently be predicted in an individual patient. This begets a 'trial and error' based approach to therapeutic regimen choice (mono or combinatorial therapeutics).

The ability to predict responsiveness to MTX, and/or other RA drugs, in an individual patient would be an invaluable clinical tool, given that response to first-line treatment is the most significant predictor of long-term outcome^{9,10}.

5

Summary of the Invention

The inventors have investigated the use of epigenetic chromosome interactions as the basis of or for use in conjunction with companion diagnostics to rheumatoid arthritis (RA), and in particular in the detection of epigenetic states to determine responsiveness to RA therapy, in particular pharmaceutical therapy of RA such as methotrexate. The inventors' work shows the role played by epigenetic interactions and provides methods for identifying the relevant chromosomal interactions. The invention relates to using chromosome interactions as the basis for companion diagnostic tests.

10

Accordingly, a **first aspect** of the present invention provides a method of determining responsiveness to a specific therapy (in particular a specific pharmaceutical therapy) for rheumatoid arthritis in a subject (preferably a mammalian such as human subject), comprising detecting the presence or absence of 5 or more (in particular 7 or more, or 10 or more, or 15 or more, or 20 or more) chromosomal interactions, wherein said chromosomal interactions are preferably at 5 or more (for example 5) different loci.

15

20

Preferably, in all aspects of the invention, said detecting comprises determining for each interaction whether or not the regions of a chromosome which are part of the interaction have been brought together.

25

More preferably, in all aspects of the invention, said detecting comprises determining for each interaction whether or not the regions of a chromosome which are part of the interaction have been brought together, by cross-linking chromosome interactions in a sample from the subject and detecting whether a sequence from both chromosome regions which are brought together is present in the cross-linked product.

30

Preferably, in all aspects of the invention, the chromosome interactions are or have been identified in an assay method that identifies chromosome interactions which are relevant to subgroups that comprises contacting a first set of nucleic acids from the subgroups with a second set of nucleic acids representing an index population of chromosome interactions, and allowing complementary sequences to hybridise, wherein the nucleic acids in the first and second sets of nucleic acids

represent (in particular are in the form of) a ligated product comprising sequences from both of the chromosome regions that have come together in the epigenetic chromosome interaction, and wherein the pattern of hybridisation between the first and second set of nucleic acids allows a determination of which epigenetic chromosome interactions are specific to subgroups in the population, wherein the subgroups differ in responsiveness to a specific therapy for rheumatoid arthritis.

Preferably, in all aspects of the invention, the feature "... the nucleic acids in the first and second sets of nucleic acids represent a ligated product comprising sequences from both of the chromosome regions that have come together in the epigenetic chromosome interaction ..." comprises or is: "... the nucleic acids in the first and second sets of nucleic acids are in the form of a ligated product(s) (preferably a ligated nucleic acid(s), more preferably ligated DNA) comprising sequences from both of the chromosome regions that have come together in the epigenetic chromosome interaction".

More preferably, in all aspects of the invention:

- the first set of nucleic acids is from at least 8 individuals, and/or
- the first set of nucleic acids is from at least 4 individuals from a first subgroup and at least 4 individuals from a second subgroup which is preferably non-overlapping with the first subgroup, and/or
- the second set of nucleic acids represents an unselected group of chromosome interactions, and/or
- the second set of nucleic acids is bound to an array at defined locations, and/or
- the second set of nucleic acids represents chromosome interactions in least 100 different genes or loci, and/or
- the second set of nucleic acids comprises at least 1000 different nucleic acids representing at least 1000 different epigenetic chromosome interactions, and/or
- the first set of nucleic acids and the second set of nucleic acids comprise nucleic acid sequences of length 10 to 100 nucleotide bases, and/or
- the first set of nucleic acids is or has been generated in a method comprising the steps: -

- (i) *in vitro* cross-linking of chromosome regions which have come together in a chromosome interaction;
- (ii) subjecting said cross-linked DNA to restriction digestion cleavage with an enzyme; and
- (iii) ligating said cross-linked cleaved DNA ends to form the first set of nucleic acids (in particular comprising ligated DNA).

Preferably, in all aspects of the invention, the subject is human; and/or the subgroups are subgroups in a human population.

5 Preferably, in all aspects of the invention:

- a. said locus is a gene, and/or
- b. a microRNA (miRNA) is expressed from the locus, and/or
- c. a non-coding RNA (ncRNA) is expressed from the locus, and/or
- d. the locus expresses a nucleic acid sequence encoding at least 10 contiguous
- 10 amino acid residues, and/or
- e. the locus expresses a regulating element.

Preferably, in all aspects of the invention, 5 or more (in particular 5 to 20, 5 to 100, 5 to 300, or 5 to 500), 7 or more (e.g. 7 to 500 or 7 to 100), more preferably 10 or more or 15 or more (e.g. 10 to 500

15 or 10 to 100 or 15 to 500 or 15 to 100), or even more preferably 20 or more (e.g. 20 to 500, 20 to 300 or 20 to 100), yet more preferably 50 or more e.g. 50 to 100, epigenetic chromosome interactions are typed.

Preferably, in the first aspect (and/or all other aspects) of the invention, the specific therapy (in

20 particular the specific pharmaceutical therapy) for rheumatoid arthritis, and/or the therapeutic agent (in particular the pharmaceutical therapeutic agent), comprises a pharmaceutically active agent (e.g. a compound or a biologic / biological agent such as a protein or antibody) suitable for use (in particular human use) in the treatment and/or prophylaxis of rheumatoid arthritis (RA), preferably a disease modifying anti-rheumatic drug (DMARD); in particular in a mammal, more particularly in a human.

25

More preferably, in all aspects of the invention, the pharmaceutically active agent comprises:

- a synthetic disease modifying anti-rheumatic drug (sDMARD), preferably comprising:

- a sDMARD which inhibits the metabolism and/or action of folic acid (preferably a sDMARD being an inhibitor of mammalian dihydrofolate reductase (DHFR), most preferably methotrexate, or

30 less preferably pemetrexed);

- sulfasalazine, or 5-aminosalicylic acid (5-ASA, mesalazine) which is an active metabolite of sulfasalazine,

- a sDMARD which is a pyrimidine synthesis inhibitor (in particular a dihydroorotate dehydrogenase (DHODH) inhibitor), most preferably leflunomide or its active metabolite teriflunomide,

- a quinolone-class antimalarial drug and sDMARD, most preferably hydroxychloroquine,

5 - a janus kinase (JAK) inhibitor sDMARD, preferably a JAK-1 and/or JAK-3 inhibitor sDMARD, most preferably tofacitinib,

- or a combination of 2, 3 or more of the sDMARDs listed herein (such a combination can, in particular, comprise or be: methotrexate + sulfasalazine, methotrexate + leflunomide, methotrexate + hydroxychloroquine, sulfasalazine + leflunomide, methotrexate + sulfasalazine +
10 hydroxychloroquine, [sulfasalazine and/or leflunomide] + hydroxychloroquine, or tofacitinib + [methotrexate, sulfasalazine, leflunomide, and/or hydroxychloroquine]);

wherein each sDMARD compound mentioned hereinabove can, independently, be in the form of the free compound and/or a pharmaceutically acceptable salt thereof; and/or

15 - a TNF-alpha (tumor necrosis factor alpha) inhibitor, in particular: a monoclonal antibody TNF-alpha inhibitor such as infliximab, adalimumab, certolizumab pegol, golimumab, or a biosimilar (in particular a USA- (e.g. FDA-) and/or European- (e.g. EMEA-) approved biosimilar) of any of these (in particular a biosimilar of infliximab such as CT-P13); and/or a circulating receptor fusion protein TNF-alpha inhibitor such as etanercept or a biosimilar thereof (in particular a USA- (e.g. FDA-) and/or European-
20 (e.g. EMEA-) approved biosimilar thereof); and/or

- a T cell costimulation inhibitor such as abatacept; and/or

- an interleukin 1 (IL-1) inhibitor such as anakinra; and/or

- a monoclonal antibody against B cells such as rituximab or a biosimilar thereof (in particular a USA- (e.g. FDA-) and/or European- (e.g. EMEA-) approved biosimilar thereof), and/or

25 - an interleukin-6 (IL-6) receptor inhibitor monoclonal antibody such as tocilizumab or a biosimilar thereof (in particular a USA- (e.g. FDA-) and/or European- (e.g. EMEA-) approved biosimilar thereof); and/or

- a glucocorticoid drug suitable for use in the treatment and/or prophylaxis of rheumatoid arthritis such as prednisone, prednisolone or dexamethasone (in particular a glucocorticoid drug in
30 combination with an sDMARD, e.g. as listed hereinabove).

Even more preferably, in all aspects of the invention, the pharmaceutically active agent comprises:

- a synthetic disease modifying anti-rheumatic drug (sDMARD), preferably comprising:

- a sDMARD which inhibits the metabolism and/or action of folic acid (preferably a sDMARD being an inhibitor of mammalian dihydrofolate reductase (DHFR), most preferably methotrexate, or less preferably pemetrexed);

- sulfasalazine, or 5-aminosalicylic acid (5-ASA, mesalazine) which is an active metabolite of sulfasalazine,

- a sDMARD which is a pyrimidine synthesis inhibitor (in particular a dihydroorotate dehydrogenase (DHODH) inhibitor), most preferably leflunomide or its active metabolite teriflunomide,

- a quinolone-class antimalarial drug and sDMARD, most preferably hydroxychloroquine,

- a janus kinase (JAK) inhibitor sDMARD, preferably a JAK-1 and/or JAK-3 inhibitor sDMARD, most preferably tofacitinib,

- or a combination of 2, 3 or more of the sDMARDs listed herein (such a combination can, in particular, comprise or be: methotrexate + sulfasalazine, methotrexate + leflunomide, methotrexate + hydroxychloroquine, sulfasalazine + leflunomide, methotrexate + sulfasalazine + hydroxychloroquine, [sulfasalazine and/or leflunomide] + hydroxychloroquine, or tofacitinib + [methotrexate, sulfasalazine, leflunomide, and/or hydroxychloroquine]);

wherein each sDMARD compound mentioned hereinabove can, independently, be in the form of the free compound and/or a pharmaceutically acceptable salt thereof.

Most preferably, in all aspects of the invention, the pharmaceutically active agent comprises:

- a synthetic disease modifying anti-rheumatic drug (sDMARD) which inhibits the metabolism and/or action of folic acid (preferably a sDMARD being an inhibitor of mammalian dihydrofolate reductase (DHFR), most preferably methotrexate, or less preferably pemetrexed);

- or a combination of methotrexate or pemetrexed (preferably methotrexate) with 1 or more of the following sDMARDs: sulfasalazine, leflunomide, hydroxychloroquine, and/or tofacitinib;

wherein each sDMARD compound mentioned hereinabove can, independently, be in the form of the free compound and/or a pharmaceutically acceptable salt thereof.

Most preferably, in all aspects of the invention, the specific therapy for rheumatoid arthritis

comprises methotrexate or a pharmaceutically acceptable salt thereof, in particular for use in the treatment and/or prophylaxis of rheumatoid arthritis.

A **second aspect** of the present invention provides an agent (in particular a pharmaceutically active agent, preferably methotrexate or a pharmaceutically acceptable salt thereof) which is therapeutic

for rheumatoid arthritis, for use in treatment and/or prophylaxis of rheumatoid arthritis in an individual (preferably in a human individual) that has been identified as being in need of said agent by a method according to the first aspect of the invention. The second aspect of the invention also provides the use of an agent (in particular a pharmaceutically active agent, preferably methotrexate or a pharmaceutically acceptable salt thereof) which is therapeutic for rheumatoid arthritis, in the manufacture of a medicament (e.g. pharmaceutical composition comprising the pharmaceutically active agent) for use in treatment and/or prophylaxis of rheumatoid arthritis in an individual (preferably in a human individual) that has been identified as being in need of said agent by a method according to the first aspect of the invention. The second aspect of the invention also provides a method of treatment and/or prophylaxis of rheumatoid arthritis in an individual (preferably in a human individual), comprising administering to the individual an agent (in particular a pharmaceutically active agent, preferably methotrexate or a pharmaceutically acceptable salt thereof) which is therapeutic for rheumatoid arthritis, wherein the individual has been identified as being in need of said agent by a method according to the first aspect of the invention.

A **third aspect** of the present invention provides a method of identifying a substance which is capable of changing in an individual (preferably in a human individual) a non-responsive state to a responsive state, in respect of the individual's responsiveness to a therapeutic agent for rheumatoid arthritis, comprising determining whether or not a candidate agent is capable of changing the chromosomal interactions from those corresponding to a non-responsive state to those which correspond to a responsive state.

A **fourth aspect** of the present invention provides a method of determining whether a candidate substance (in particular a pharmaceutically active agent) is suitable for the treatment and/or prophylaxis of rheumatoid arthritis, comprising detecting the change in epigenetic chromosome interactions caused by the drug (i.e. the candidate substance, in particular a pharmaceutically active agent), wherein said interactions relate to the mechanism of action of the drug or the pharmacodynamics properties of the drug.

A **fifth aspect** of the present invention provides a library of nucleic acids (e.g. DNA and/or isolated nucleic acids) which comprises at least 200 different second nucleic acids (e.g. DNA and/or isolated nucleic acids), as defined herein, optionally bound to an array.

Preferably, in the fifth aspect of the invention, the library comprises 5 or more, 7 or more, 10 or more, 15 or more, 20 or more, 25 or more, 30 or more, or 40 or more, or 50 or more, nucleic acids (e.g. DNA and/or isolated nucleic acids) each of which comprise (for example each of which consist essentially of, e.g. consist of) a nucleic acid sequence (e.g. DNA sequence) selected from the group consisting of:

(i) the nucleic acid (e.g. DNA) sequences listed in Table 7a and/or Table 8a and/or Table 9 (preferably Table 7a and/or Table 8a, most preferably Table 7a); and

(ii) nucleic acid (e.g. DNA) sequences having at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% identity to one or more sequences listed in Table 7a and/or Table 8a and/or Table 9 (preferably Table 7a and/or Table 8a, most preferably Table 7a).

A **sixth aspect** of the present invention provides a library of nucleic acids (e.g. DNA and/or isolated nucleic acids) comprising 5 or more, 7 or more, 10 or more, 15 or more, 20 or more, 25 or more, 30 or more, or 40 or more, or 50 or more, or 70 or more, nucleic acids (e.g. DNA and/or isolated nucleic acids), each of which comprise (for example each of which consist essentially of, e.g. consist of) a nucleic acid sequence (e.g. DNA sequence) selected from the group consisting of:

(i) the nucleic acid (e.g. DNA) sequences listed in Table 7a and/or Table 8a and/or Table 9 (preferably Table 7a and/or Table 8a, most preferably Table 7a); and

(ii) nucleic acid (e.g. DNA) sequences having at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% identity to one or more sequences listed in Table 7a and/or Table 8a and/or Table 9 (preferably Table 7a and/or Table 8a, most preferably Table 7a).

A **seventh aspect** of the present invention provides a nucleic acid (e.g. DNA and/or an isolated nucleic acid) comprising (for example consisting essentially of, e.g. consisting of) a nucleic acid sequence (e.g. DNA sequence) selected from the group consisting of:

(i) the nucleic acid (e.g. DNA) sequences listed in Table 7a and/or Table 8a and/or Table 9 (preferably Table 7a and/or Table 8, most preferably Table 7a); and

(ii) nucleic acid (e.g. DNA) sequences having at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% identity to one or more sequences listed in Table 7a and/or Table 8a and/or Table 9 (preferably Table 7a and/or Table 8a, most preferably Table 7a).

The invention also provides a nucleic acid (e.g. DNA and/or an isolated nucleic acid) comprising (for example consisting essentially of, e.g. consisting of) a nucleic acid sequence (e.g. DNA sequence) selected from the nucleic acid (e.g. DNA) sequences listed in Table 7a and/or Table 8a and/or Table 9 (preferably Table 7a and/or Table 8a, most preferably Table 7a).

5

For clarity, sequence identity is the amount of nucleotide characters that match exactly between two sequences, and these values are typically estimated by common algorithms such as BLAST and/or BLAT. See hereinafter under "Homologues" for more information.

10

Brief Description of the Drawings

Figure 1 is a figure comprising pie-charts and graphs relating to: Chromosome Conformation Signature EpiSwitch™ Markers discriminate MTX responders (R) from non-responders (NR). A discovery cohort of responder (R) and non-responder (NR) RA patients were selected based on DAS28 (Disease Activity Score of 28 joints) EULAR (The European League Against Rheumatism) response criteria (see methods). (A) Pie charts show the clinical interpretation of CDAI scores for both R and NR patients at baseline and 6 months. (B) CDAI scores of R and NR patients at baseline and 6 months. (C) EpiSwitch™ array analysis of peripheral blood mononuclear cells taken at diagnosis from R and NR, and healthy controls (HC) identified 922 statistically significant stratifying marker candidates. Further analysis revealed that 420 were specific for NR, 210 to R and 159 to HC. Pie charts show the proportion in relation to the 13,322 conditional chromosome conformations screened. All markers showed adjusted $p < 0.2$. (D) Hierarchical clustering using Manhattan distance measure with complete linkage agglomeration is shown by the heatmaps. Marker selection using binary patterning across the 3 groups (R, NR and HC) initially reduced the 922 EpiSwitch™ Markers to 65 and then the top 30 markers.

25

Figure 2 is a figure comprising pie-charts and graphs relating to: Refinement and validation of the Chromosome Conformation Signature EpiSwitch™ Markers. The validation cohort of responder (R) and non-responder (NR) RA patients were selected based on DAS28 (Disease Activity Score of 28 joints) EULAR (The European League Against Rheumatism) response criteria (see methods). (A) Pie charts show the clinical interpretation of CDAI scores for both R and NR patients at baseline and 6 months. (B) CDAI scores of R and NR patients at baseline and 6 months. **** $P < 0.0001$ by Kruskal–Wallis test with Dunn's multiple comparison post test (C) Correlation plot of the classifying 5 EpiSwitch™ markers. The red box indicates the markers that define NR whilst the orange box indicated

30

markers that define R. (D) Principle Component Analysis (PCA) for a 60 patient cohort based on their binary scores for the classifying 5 EpiSwitch™ markers.

Figure 3 is a figure comprising graphs relating to: Prognostic stratification and model validation for response to methotrexate (MTX) treatment. (A) Representative examples of 5 selected Receiver Operating Characteristics (ROC) curves from 150 randomisations of the data using the 5 CCS marker logistic regression classifiers. (B) Factor Analysis for responder (R) and non-responder (NR) RA patients vs healthy controls (HC) using EpiSwitch™ CCS markers selected for discerning MTX responders from MTX non-responders.

Figure 4 is a Schematic diagram of the 3C extraction process. 3C means chromatin conformation capture, or chromosome conformation capture.

Figure 5 is a Scheme illustrating the Design for Discovery and Validation of Epigenetic Stratifying Biomarker Signature for DMARDS Naïve ERA patients, who were confirmed within 6 months of MTX treatment as responders (N) or non-responders (NR). Epigenetic stratification was based on conditional chromosome confirmations screened and monitored by EpiSwitch™ Array and PCR (polymerase chain reaction) platforms. Disease specific epigenetic nature of the identified biomarkers was confirmed by stratification against healthy controls (HC). Validation was performed on 60 RA patients (30 responders and 30 non-responders) and 30 HC.

Detailed Description of the Invention

The invention has several different aspects:

- a method of determining responsiveness to a specific therapy for rheumatoid arthritis in a subject;
- a companion diagnostic method;
- a therapeutic agent for use in treatment and/or prophylaxis of an individual (specifically, in the treatment and/or prophylaxis of rheumatoid arthritis in an individual, in particular in a human individual), wherein said individual has been identified as being in need of the therapeutic agent in particular by a method of determining responsiveness and/or a companion diagnostic method of the invention;
- a method of screening for (identifying) an agent, in particular a therapeutic agent, which is capable of changing responsiveness (in particular of an individual e.g. human individual) to a therapy for rheumatoid arthritis;

- a method of determining the effect of a drug (e.g. therapeutic agent) comprising detecting the change in epigenetic chromosome interactions caused by the drug.

Epigenetic Interactions

As used herein, the term 'epigenetic' interactions typically refers to interactions between distal regions of a locus on a chromosome, said interactions being dynamic and altering, forming or breaking depending upon the status of the region of the chromosome.

In particular methods of the invention chromosome interactions are detected by first generating a ligated nucleic acid that comprises sequence(s) from both regions of the chromosomes that are part of the chromosome interactions. In such methods the regions can be cross-linked by any suitable means. In a preferred embodiment, the interactions are cross-linked using formaldehyde, but may also be cross-linked by any aldehyde, or D-Biotinoyl-e- aminocaproic acid-N-hydroxysuccinimide ester or Digoxigenin-3-O-methylcarbonyl- e-aminocaproic acid-N-hydroxysuccinimide ester. Para-formaldehyde can cross link DNA chains which are 4 Angstroms apart.

The chromosome interaction may reflect the status of the region of the chromosome, for example, if it is being transcribed or repressed in response to change of the physiological conditions.

Chromosome interactions which are specific to subgroups as defined herein have been found to be stable, thus providing a reliable means of measuring the differences between the two subgroups.

In addition, chromosome interactions specific to a disease condition will normally occur early in the disease process, for example compared to other epigenetic markers such as methylation or changes to binding of histone proteins. Thus the companion diagnostic method of the invention is able to detect early stages of a disease state. This allows early treatment which may as a consequence be more effective. Another advantage of the invention is that no prior knowledge is needed about which loci are relevant for identification of relevant chromosome interactions. Furthermore there is little variation in the relevant chromosome interactions between individuals within the same subgroup. Detecting chromosome interactions is highly informative with up to 50 different possible interactions per gene, and so methods of the invention can interrogate 500,000 different interactions.

Location and Causes of Epigenetic Interactions

Epigenetic chromosomal interactions may overlap and include the regions of chromosomes shown to encode relevant or undescribed genes, but equally may be in intergenic regions. It should further be noted that the inventors have discovered that epigenetic interactions in all regions are equally

important in determining the status of the chromosomal locus. These interactions are not necessarily in the coding region of a particular gene located at the locus and may be in intergenic regions.

- 5 The chromosome interactions which are detected in the invention could be caused by changes to the underlying DNA sequence, by environmental factors, DNA methylation, non-coding antisense RNA transcripts, non-mutagenic carcinogens, histone modifications, chromatin remodelling and specific local DNA interactions. The changes which lead to the chromosome interactions may be caused by changes to the underlying nucleic acid sequence, which themselves do not directly affect
- 10 a gene product or the mode of gene expression. Such changes may be for example, SNP's within and/or outside of the genes, gene fusions and/or deletions of intergenic DNA, microRNA, and non-coding RNA. For example, it is known that roughly 20% SNPs are in non-coding regions, and therefore the method as described is also informative in non-coding situation. In one embodiment the regions of the chromosome which come together to form the interaction are less than 5 kb, 3 kb,
- 15 1 kb, 500 base pairs or 200 base pairs apart.

The chromosome interaction which is detected in the companion diagnostic method is preferably one which is within any of the genes mentioned in the Tables herein. However it may also be upstream or downstream of the genes, for example up to 50,000, 30,000, 20,000, 10,000 or 5000 bases upstream or downstream from the gene or from the coding sequence.

- 20 The chromosome interaction which is detected may or may not be one which occurs between a gene (including coding sequence) and its regulatory region, such as a promoter. The chromosome interaction which is detected may or may not be one which is inherited, for example an inherited imprinted characteristic of a gene region. The individual may be male or female. The individual may be 30 years old or older. The individual may be 29 years old or younger.

25

Types of Clinical Situation

The specific case of use of methotrexate (MTX) to treat RA (Rheumatoid Arthritis) illustrates the general principles. There are currently no tests that clinicians can use *a priori* to determine if patients will respond to MTX when the patients are first given the drug. Since a significant number

30 (about 30%) of patients do not respond to MTX, being able to predict whether a patient is a responder or non-responder will increase the chances of successfully treating RA, as well as saving time and money.

The invention allows stratification based on biomarkers for specific phenotypes relating to rheumatoid arthritis, i.e. by recognising a particular chromosome confirmation signature and/or a change in that particular signature.

The method may or may not be used for diagnosis of the presence of rheumatoid arthritis. The methods of the invention can be used to type loci where the mechanisms of disease are unknown, unclear or complex. Detection of chromosome interactions provides an efficient way of following changes at the different levels of regulation, some of which are complex. For example in some cases around 37,000 non-coding RNAs can be activated by a single impulse.

Subgroups and Personalised Treatment

As used herein, a “subgroup” preferably refers to a population subgroup (a subgroup in a population), more preferably a subgroup in a or the population of a particular animal such as a particular mammal (e.g. human, non-human primate, or rodent e.g. mouse or rat) or a particular nematode worm (e.g. *C. elegans*). Most preferably, a “subgroup” refers to a subgroup in a or the human population.

Particular populations, e.g. human populations, of interest include: the human population overall, the human RA population (i.e. humans suffering from RA), the human healthy population (healthy controls), the human population which is healthy in the sense of not suffering from RA, the human (healthy and/or RA) population who are responders to a particular drug / therapy, or the human (healthy and/or RA) population who are non-responders to a particular drug / therapy.

The invention relates to detecting and treating particular subgroups in a population, preferably in a or the human population. Within such subgroups the characteristics discussed herein (such as responsiveness to treatment and/or prophylaxis; in particular responsiveness to a specific e.g. pharmaceutical treatment and/or prophylaxis e.g. to a therapeutically active substance / therapeutic agent e.g. pharmaceutical therapeutic agent) will be present or absent. Epigenetic interaction differences on a chromosome are, generally speaking, structural differences which exist at a genomic level. The inventors have discovered that these differ between subsets (for example two, or two or more, subsets) in a given population. Identifying these differences will allow physicians to categorize their patients as a part of one subset of the population as described in the method. The invention therefore provides physicians with a method of personalizing medicine for the patient based on their epigenetic chromosome interactions, and provide an alternative more effective treatment and/or prophylaxis regime.

In another embodiment, threshold levels for determining to what extent a subject is defined as belonging to one subgroup and not to a or the other subgroup of the population (e.g. human

population, e.g. human RA population) are applied. In one preferable embodiment wherein the subgroups comprise responders versus non-responders of a therapy for the treatment of a particular disease (e.g. or i.e. RA), said threshold may be measured by change in DAS28 score (Disease Activity Score of 28 joints). In one embodiment, a score above 1.2 units indicates a subject falls into the responder subgroup, whilst a score below 1.2 units indicates a subject is defined as a non-responder.

Typically a subgroup will be at least 10%, at least 30%, at least 50%, at least 70%, or at least 80% of the general population.

Generating Ligated Nucleic Acids

Certain embodiments of the invention utilise ligated nucleic acids, in particular ligated DNA. These comprise sequences from both of the regions that come together in a chromosome interaction and therefore provide information about the interaction. The EpiSwitch™ method described herein uses generation of such ligated nucleic acids to detect chromosome interactions.

One such method, in particular one particular method of detecting chromosome interactions and/or one particular method of determining epigenetic chromosome interactions and/or one particular method of generating ligated nucleic acids (e.g. DNA), comprises the steps of:

- (i) *in vitro* crosslinking of said epigenetic chromosomal interactions present at the chromosomal locus;
- (ii) optionally isolating the cross-linked DNA from said chromosomal locus;
- (iii) subjecting said cross-linked DNA to restriction digestion with an enzyme that cuts it at least once (in particular an enzyme that cuts at least once within said chromosomal locus);
- (iv) ligating said cross-linked cleaved DNA ends (in particular to form DNA loops); and
- (v) identifying the presence of said ligated DNA and/or said DNA loops, in particular using techniques such as PCR (polymerase chain reaction), to identify the presence of a specific chromosomal interaction.

One particularly preferable method of detecting, determining and/or monitoring chromosome interactions and/or epigenetic changes, involving *inter alia* the above-mentioned steps of crosslinking, restriction digestion, ligating, and identifying, is disclosed in WO 2009/147386 A1 (Oxford Biodynamics Ltd), the entire disclosure of which (in particular claims 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or 21 of which) are incorporated herein by reference as though fully set forth. Claim 1 of WO 2009/147386 A1, which can be used in those methods of the

present invention which involve a ligated product(s) and/or a ligated nucleic acid(s), discloses a method of monitoring epigenetic changes comprising monitoring changes in conditional long range chromosomal interactions at at least one chromosomal locus where the spectrum of long range interaction is associated with a specific physiological condition, said method comprising the steps

5 of:-

- (i) *in vitro* crosslinking of said long range chromosomal interactions present at the at least one chromosomal locus;
- (ii) isolating the cross linked DNA from said chromosomal locus;
- (iii) subjecting said cross linked DNA to restriction digestion with an enzyme that cuts at least once
- 10 within the at least one chromosomal locus;
- (iv) ligating said cross linked cleaved DNA ends to form DNA loops; and
- (v) identifying the presence of said DNA loops;

wherein the presence of DNA loops indicates the presence of a specific long range chromosomal interaction.

- 15 PCR (polymerase chain reaction) may be used to detect or identify the ligated nucleic acid, for example the size of the PCR product produced may be indicative of the specific chromosome interaction which is present, and may therefore be used to identify the status of the locus. The skilled person will be aware of numerous restriction enzymes which can be used to cut the DNA within the chromosomal locus of interest. It will be apparent that the particular enzyme used will
- 20 depend upon the locus studied and the sequence of the DNA located therein. A non-limiting example of a restriction enzyme which can be used to cut the DNA as described in the present invention is TakaRa LA Taq polymerase.

Embodiments such as EpiSwitch™ Technology

- The EpiSwitch™ Technology relates to the use of microarray EpiSwitch™ marker data in the
- 25 detection of epigenetic chromosome conformation signatures specific for phenotypes. The present inventors describe herein how the EpiSwitch™ Array Platform has been used for discovery of chromosome signature pool of potential biomarkers specific for particular disadvantageous phenotypes subgroups versus healthy controls. The inventors also provide examples of validated use and translation of chromosome conformation signatures from microarray into PCR platform with
- 30 examples of several markers specific between subgroups from the cohorts tested on the array.

Embodiments such as EpiSwitch™ which utilise ligated nucleic acids in the manner described herein (for identifying relevant chromosome interactions and in companion diagnostic methods) have

several advantages. They have a low level of stochastic noise, for example because the nucleic acid sequences from the first set of nucleic acids of the present invention either hybridise or fail to hybridise with the second set of nucleic acids. This provides a binary result permitting a relatively simple way to measure a complex mechanism at the epigenetic level. EpiSwitch™ technology also has fast processing time and low cost. In one embodiment the processing time is 3 hours to 6 hours.

Samples and Sample Treatment

The sample will contain DNA from the individual. It will normally contain cells. In one embodiment a sample is obtained by minimally invasive means, and may for example be blood. DNA may be extracted and cut up with standard restriction enzymes. This can pre-determine which chromosome conformations are retained and will be detected with the EpiSwitch™ platforms. In one embodiment wherein the sample is a blood sample previously obtained from the patient, the described method is advantageous because the procedure is minimally invasive. Due to the synchronisation of chromosome interactions between tissues and blood, including horizontal transfer, a blood sample can be used to detect the chromosome interactions in tissues, such as tissues relevant to disease. For certain conditions, such as cancer, genetic noise due to mutations can affect the chromosome interaction 'signal' in the relevant tissues and therefore using blood is advantageous.

Properties of Nucleic acids of the invention

The disclosure herein mentions first and second nucleic acids. In addition the nucleic acids are used in the companion diagnostic method and in other embodiments to detect the presence or absence of chromosome interactions (for example by binding to ligated nucleic acids generated from samples). The nucleic acids of the invention typically comprise two portions each comprising sequence from one of the two regions of the chromosome which come together in the chromosome interaction. Typically each portion is at least 8, 10, 15, 20, 30 or 40 nucleotides in length. Preferred nucleic acids comprise sequence from any of the genes mentioned in the tables, in particular where the nucleic acid is used in embodiments relevant to the condition relevant for that table. Preferred nucleic acids comprise the specific probe sequences mentioned in the tables for specific conditions or fragments or homologues of such sequences. Preferably the nucleic acids are DNA. It is understood that where a specific sequence is provided the invention may use the complementary as required in the particular embodiment.

The Second Set of Nucleic Acids – the ‘Index’ Sequences

The second set of nucleic acid sequences has the function of being an index, and is essentially a set of nucleic acid sequences which are suitable for identifying subgroup specific sequence. They can represent the ‘background’ chromosomal interactions and might be selected in some way or be
5 unselected. They are a subset of all possible chromosomal interactions.

The second set of nucleic acids may be derived by any suitable method. They can be derived computationally or they may be based on chromosome interaction in individuals. They typically represent a larger population group than the first set of nucleic acids. In one embodiment, the second set of nucleic acids represents all possible epigenetic chromosomal interactions in a specific
10 set of genes. In another embodiment, the second set of nucleic acids represents a large proportion of all possible epigenetic chromosomal interactions present in a population described herein. In one embodiment, the second set of nucleic acids represents at least 50% or at least 80% of epigenetic chromosomal interactions in at least 20, 50, 100 or 500 genes.

The second set of nucleic acids typically represents at least 100 possible epigenetic chromosome
15 interactions which modify, regulate or any way mediate a disease state / phenotype in population. The second set of nucleic acids may represent chromosome interactions that affect a disease state in a species, for example comprising nucleic acid sequences which encode cytokines, kinases, or regulators associated with any disease state, predisposition to a disease or a disease phenotype. The second set of nucleic acids comprises sequences representing epigenetic interactions relevant and
20 not relevant to the companion diagnostic method.

In one embodiment the second set of nucleic acids derive at least partially from naturally occurring sequences in a population, and are typically obtained by *in silico* methods. Said nucleic acids may further comprise single or multiple mutations in comparison to a corresponding portion of nucleic acids present in the naturally occurring nucleic acids. Mutations include deletions, substitutions
25 and/or additions of one or more nucleotide base pairs. In one embodiment, the second set of nucleic acids may comprise sequence representing a homologue and/or orthologue with at least 70% sequence identity to the corresponding portion of nucleic acids present in the naturally occurring species. In another embodiment, at least 80% sequence identity or at least 90% sequence identity to the corresponding portion of nucleic acids present in the naturally occurring species is
30 provided.

Properties of the second set of nucleic acids

In one embodiment, there are at least 100 different nucleic acid sequences in the second set of nucleic acids, preferably at least 1000, 2000 or 5000 different nucleic acids sequences, with up to
5 100,000, 1,000,000 or 10,000,000 different nucleic acid sequences. A typical number would be 100 to 1,000,000, such as 1,000 to 100,000 different nucleic acids sequences. All or at least 90% or at least 50% or these would correspond to different chromosomal interactions.

In one embodiment, the second set of nucleic acids represent chromosome interactions in at least 20 different loci or genes, preferably at least 40 different loci or genes, and more preferably at least
10 100, at least 500, at least 1000 or at least 5000 different loci or genes, such as 100 to 10,000 different loci or genes.

The lengths of the second set of nucleic acids are suitable for them to specifically hybridise according to Watson Crick base pairing to the first set of nucleic acids to allow identification of chromosome interactions specific to subgroups. Typically the second set of nucleic acids will comprise two
15 portions corresponding in sequence to the two chromosome regions which come together in the chromosome interaction. The second set of nucleic acids typically comprise nucleic acid sequences which are at least 10, preferably 20, and preferably still 30 bases (nucleotides) in length. In another embodiment, the nucleic acid sequences may be at the most 500, preferably at most 100, and preferably still at most 50 base pairs in length. In a preferred embodiment, the second set of nucleic
20 acids comprises nucleic acid sequences of between 17 and 25 base pairs. In one embodiment at least 100, 80% or 50% of the second set of nucleic acid sequences have lengths as described above. Preferably the different nucleic acids do not have any overlapping sequences, for example at least 100%, 90%, 80% or 50% of the nucleic acids do not have the same sequence over at least 5 contiguous nucleotides.

Given that the second set of nucleic acids acts as an 'index' then the same set of second nucleic acids
25 may be used with different sets of first nucleic acids which represent subgroups for different characteristics, i.e. the second set of nucleic acids may represent a 'universal' collection of nucleic acids which can be used to identify chromosome interactions relevant to different disease characteristics.

The first set of nucleic acids

The first set of nucleic acids are normally from individuals known to be in two or more distinct subgroups defined by presence or absence of a characteristic relevant to a companion diagnostic, such as any such characteristic mentioned herein. The first nucleic acids may have any of the characteristics and properties of the second set of nucleic acids mentioned herein. The first set of nucleic acids is normally derived from a sample from the individuals which has undergone treatment and processing as described herein, particularly the EpiSwitch™ cross-linking and cleaving steps. Typically the first set of nucleic acids represents all or at least 80% or 50% of the chromosome interactions present in the samples taken from the individuals.

Typically, the first set of nucleic acids represents a smaller population of chromosome interactions across the loci or genes represented by the second set of nucleic acids in comparison to the chromosome interactions represented by second set of nucleic acids, i.e. the second set of nucleic acids is representing a background or index set of interactions in a defined set of loci or genes.

Library of Nucleic Acids

The invention provides a library of nucleic acids which comprises at least 200, 500, 1000, 5000 or at least 10,000 different nucleic acids from the second set of nucleic acids. The invention provides a particular library of nucleic acids which typically comprises at least 200 different nucleic acids. The library of nucleic acids may have any of the characteristics or properties of the second set of nucleic acids mentioned herein. The library may be in the form of nucleic acids bound to an array.

Hybridisation

The invention requires a means for allowing wholly or partially complementary nucleic acid sequences from the first set of nucleic acids and the second set of nucleic acids to hybridise. In one embodiment all of the first set of nucleic acids is contacted with all of the second set of nucleic acids in a single assay, i.e. in a single hybridisation step. However any suitable assay can be used.

Labelled Nucleic Acids and Pattern of Hybridisation

The nucleic acids mentioned herein may be labelled, preferably using an independent label such as a fluorophore (fluorescent molecule) or radioactive label which assists detection of successful hybridisation. Certain labels can be detected under UV light.

The pattern of hybridisation, for example on an array described herein, represents differences in epigenetic chromosome interactions between the two subgroups, and thus provides a method of comparing epigenetic chromosome interactions and determination of which epigenetic chromosome interactions are specific to a subgroup in the population of the present invention.

- 5 The term 'pattern of hybridisation' broadly covers the presence and absence of hybridisation between the first and second set of nucleic acids, i.e. which specific nucleic acids from the first set hybridise to which specific nucleic acids from the second set, and so it not limited to any particular assay or technique, or the need to have a surface or array on which a 'pattern' can be detected.

Companion Diagnostic Method

- 10 The invention provides a companion diagnostic method based on information provided by chromosome interactions. Two distinct companion diagnostic methods are provided which identify whether an individual has a particular characteristic relevant to a companion diagnostic. One method is based on typing a locus in any suitable way and the other is based on detecting the presence or absence of chromosome interactions. The characteristic may be any one of the
- 15 characteristics mentioned herein relating to a condition. The companion diagnostic method can be carried out at more than one time point, for example where monitoring of an individual is required.

Companion Diagnostic Method Based on Typing a Locus

- The method of the invention which identified chromosome interactions that are specific to subgroups can be used to identify a locus, which may be a gene that can be typed as the basis of
- 20 companion diagnostic test. Many different gene-related effects can lead to the same chromosome interaction occurring. In this embodiment any characteristic of the locus may be typed, such as presence of a polymorphism in the locus or in an expressed nucleic acid or protein, the level of expression from the locus, the physical structure of the locus or the chromosome interactions present in the locus. In one particular embodiment the locus may be any of the genes mentioned
- 25 herein in the tables, in particular in Tables 1, 3, 5, 7, 8 and/or 9 (in particular Tables 1, 3 and/or 5), or any property of a locus which is in the vicinity of a chromosome interaction found to be linked to the relevant condition.

Companion Diagnostic Method Based on Detecting Chromosome Interactions

- The invention provides a companion diagnostic method which comprises detecting the presence or
- 30 absence of chromosome interactions, typically 5 to 20 or 5 to 500 such interactions, preferably 20 to

300 or 50 to 100 interactions, in order to determine the presence or absence of a characteristic in an individual. Preferably the chromosome interactions are those in any of the genes mentioned herein. In one particular embodiment the chromosome interactions which are typed are those represented by the nucleic acids disclosed in the tables herein, in particular in in Tables 7a, 8a and/or 9, for example when the method is for the purpose of determining the presence or absence of characteristics defined in those tables.

Specific Conditions

The companion diagnostic method can be used to detect the presence of any of the specific conditions or characteristics mentioned herein. The companion diagnostic method can be used to detect responsiveness to methotrexate in rheumatoid arthritis patients.

Preferably the presence or absence of any of the chromosome interactions within any of the relevant genes mentioned in the tables is detected. For example in at least 1, 3, 10, 20, 50 of the genes mentioned in any one of the tables. Preferably the presence or absence of chromosome interactions represented by the probes sequences in the Tables I s determined in the method. For example at least 1, 3, 10, 20, 50, or 100 of the relevant chromosome interactions from any one of the tables. These numbers of genes or chromosome interactions can be used in any of the different embodiments mentioned herein.

The Individual Tested Using the Companion Diagnostic Method

The individual to be tested may or may not have any symptoms of any disease condition or characteristic mentioned herein. The individual may be at risk of any such condition or characteristic. The individual may have recovered or be in the process of recovering from the condition or characteristic. The individual is preferably a mammal, such as a primate, human or rodent.

Screening Method

A method of identifying a substance which is capable of changing in an individual a non-responsive state to a responsive state to a therapeutic agent for rheumatoid arthritis comprising determining whether a candidate agent is capable of changing the chromosomal interactions from those corresponding to a non-responsive state to those which correspond to a responsive state.

In one particular embodiment the method determines whether a candidate agent is capable of changing any chromosomal interaction mentioned herein.

The method may be carried out in vitro (inside or outside a cell) or in vivo (upon a non-human organism). In one particular embodiment the method is carried out on a cell, cell culture, cell
 5 extract, tissue, organ or organism, such as one which comprises the relevant chromosome interaction(s). The method is typically carried out by contacting (or administering) the candidate agent with the gene, cell, cell culture, cell extract, tissue, organ or organism.

Suitable candidate substances which tested in the above screening methods include antibody agents (for example, monoclonal and polyclonal antibodies, single chain antibodies, chimeric antibodies and
 10 CDR-grafted antibodies). Furthermore, combinatorial libraries, defined chemical identities, peptide and peptide mimetics, oligonucleotides and natural agent libraries, such as display libraries (e.g. phage display libraries) may also be tested. The candidate substances may be chemical compounds, which are typically derived from synthesis around small molecules which may have any of the properties of the agent mentioned herein.

15 *Preferred Loci, Genes and Chromosome Interactions*

For all aspects of the invention preferred loci, genes and chromosome interactions are mentioned in the tables. For all aspects of the invention preferred loci, genes and chromosome interactions are provided in the tables. Typically the methods chromosome interactions are detected from at least 1,
 3, 10, 20, 30 or 50 of the relevant genes listed in the table. Preferably the presence or absence of at
 20 least 1, 3, 10, 20, 30 or 50 of the relevant specific chromosome interactions represented by the probe sequences in any one table is detected.

The loci may be upstream or downstream of any of the genes mentioned herein, for example 50 kb upstream or 20 kb downstream.

Preferred Embodiments for Sample Preparation and Chromosome Interaction Detection

25 Methods of preparing samples and detecting chromosome conformations are described herein. Optimised (non-conventional) versions of these methods can be used, for example as described in this section.

Typically the sample will contain at least 2×10^5 cells. The sample may contain up to 5×10^5 cells. In one embodiment, the sample will contain 2×10^5 to 5.5×10^5 cells

30 Crosslinking of epigenetic chromosomal interactions present at the chromosomal locus is described herein. This may be performed before cell lysis takes place. Cell lysis may be performed for 3 to 7

minutes, such as 4 to 6 or about 5 minutes. In some embodiments, cell lysis is performed for at least 5 minutes and for less than 10 minutes.

Digesting DNA with a restriction enzyme is described herein. Typically, DNA restriction is performed at about 55°C to about 70°C, such as for about 65°C, for a period of about 10 to 30 minutes, such as about 20 minutes.

Preferably a frequent cutter restriction enzyme is used which results in fragments of ligated DNA with an average fragment size up to 4000 base pair. Optionally the restriction enzyme results in fragments of ligated DNA have an average fragment size of about 200 to 300 base pairs, such as about 256 base pairs. In one embodiment, the typical fragment size is from 200 base pairs to 4,000 base pairs, such as 400 to 2,000 or 500 to 1,000 base pairs.

In one embodiment of the EpiSwitch method a DNA precipitation step is not performed between the DNA restriction digest step and the DNA ligation step.

DNA ligation is described herein. Typically the DNA ligation is performed for 5 to 30 minutes, such as about 10 minutes.

The protein in the sample may be digested enzymatically, for example using a proteinase, optionally Proteinase K. The protein may be enzymatically digested for a period of about 30 minutes to 1 hour, for example for about 45 minutes. In one embodiment after digestion of the protein, for example Proteinase K digestion, there is no cross-link reversal or phenol DNA extraction step.

In one embodiment PCR detection is capable of detecting a single copy of the ligated nucleic acid, preferably with a binary read-out for presence/absence of the ligated nucleic acid.

Homologues

Homologues of polynucleotide / nucleic acid (e.g. DNA) sequences are referred to herein. Such homologues typically have at least 70% homology, preferably at least 80, at least 85%, at least 90%, at least 95%, at least 97%, at least 98% or at least 99% homology, for example over a region of at least 10, 15, 20, 30, 100 or more contiguous nucleotides, or across the portion of the nucleic acid which is from the region of the chromosome involved in the chromosome interaction. The homology may be calculated on the basis of nucleotide identity (sometimes referred to as "hard homology").

Therefore, in a particular embodiment, homologues of polynucleotide / nucleic acid (e.g. DNA) sequences are referred to herein by reference to % sequence identity. Typically such homologues have at least 70% sequence identity, preferably at least 80%, at least 85%, at least 90%, at least 95%,

at least 97%, at least 98% or at least 99% sequence identity, for example over a region of at least 10, 15, 20, 30, 100 or more contiguous nucleotides, or across the portion of the nucleic acid which is from the region of the chromosome involved in the chromosome interaction.

- 5 For example the UWGCG Package provides the BESTFIT program which can be used to calculate homology and/or % sequence identity (for example used on its default settings) (Devereux et al (1984) *Nucleic Acids Research* 12, p387-395). The PILEUP and BLAST algorithms can be used to calculate homology and/or % sequence identity and/or line up sequences (such as identifying equivalent or corresponding sequences (typically on their default settings), for example as described
10 in Altschul S. F. (1993) *J Mol Evol* 36:290-300; Altschul, S, F et al (1990) *J Mol Biol* 215:403-10.

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pair (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some
15 positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighbourhood word score threshold (Altschul et al, supra). These initial neighbourhood word hits act as seeds for initiating searches to find HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extensions for the word hits in each direction are halted when:
20 the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W5 T and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919) alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a
25 comparison of both strands.

The BLAST algorithm performs a statistical analysis of the similarity between two sequences; see e.g., Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90: 5873-5787. One measure of similarity
30 provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two polynucleotide sequences would occur by chance. For example, a sequence is considered similar to another sequence if the smallest sum probability in comparison of the first sequence to the second sequence is less than about 1, preferably less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

The homologous sequence typically differs by 1, 2, 3, 4 or more bases, such as less than 10, 15 or 20 bases (which may be substitutions, deletions or insertions of nucleotides). These changes may be measured across any of the regions mentioned above in relation to calculating homology and/or % sequence identity.

Arrays

The second set of nucleic acids may be bound to an array, and in one embodiment there are at least 15,000, 45,000, 100,000 or 250,000 different second nucleic acids bound to the array, which preferably represent at least 300, 900, 2000 or 5000 loci. In one embodiment one, or more, or all of the different populations of second nucleic acids are bound to more than one distinct region of the array, in effect repeated on the array allowing for error detection. The array may be based on an Agilent SurePrint G3 Custom CGH microarray platform. Detection of binding of first nucleic acids to the array may be performed by a dual colour system.

Therapeutic Agents

Therapeutic agents are mentioned herein. The invention provides such agents for use in preventing or treating the relevant condition. This may comprise administering to an individual in need a therapeutically effective amount of the agent. The invention provides use of the agent in the manufacture of a medicament to prevent or treat the disease. The methods of the invention may be used to select an individual for treatment. The methods of the invention, and in particular the method for carrying out a companion diagnostic test, may include a treatment step where a person identified by the method may then be administered with an agent that prevents or treats the relevant condition.

The formulation of the agent will depend upon the nature of the agent. The agent will be provided in the form of a pharmaceutical composition containing the agent and a pharmaceutically acceptable carrier or diluent. Suitable carriers and diluents include isotonic saline solutions, for example phosphate-buffered saline. Typical oral dosage compositions include tablets, capsules, liquid solutions and liquid suspensions. The agent may be formulated for parenteral, intravenous, intramuscular, subcutaneous, transdermal or oral administration.

The dose of agent may be determined according to various parameters, especially according to the

substance used; the age, weight and condition of the individual to be treated; the route of administration; and the required regimen. A physician will be able to determine the required route of administration and dosage for any particular agent. A suitable dose may however be from 0.1 to 100 mg/kg body weight such as 1 to 40 mg/kg body weight, for example, to be taken from 1 to 3 times daily.

Forms of the Substance Mentioned Herein

Any of the substances, such as nucleic acids or therapeutic agents, mentioned herein may be in purified or isolated form. The may be in a form which is different from that found in nature, for example they may be present in combination with other substance with which they do not occur in nature. The nucleic acids (including portions of sequences defined herein) may have sequences which are different to those found in nature, for example having at least 1, 2, 3, 4 or more nucleotide changes in the sequence as described in the section on homology. The nucleic acids may have heterologous sequence at the 5' or 3' end. The nucleic acids may be chemically different from those found in nature, for example they may be modified in some way, but preferably are still capable of Watson-Crick base pairing. Where appropriate the nucleic acids will be provided in double stranded or single stranded form. The invention provides all the of specific nucleic acid sequences mentioned herein in single or double stranded form, and thus includes the complementary strand to any sequence which is disclosed.

The invention also provides a kit for carrying out any method of the invention, including detection of a chromosomal interaction associated with a particular subgroup. Such a kit can include a specific binding agent capable of detecting the relevant chromosomal interaction, such as agents capable of detecting a ligated nucleic acid generated by methods of the invention. Preferred agents present in the kit include probes capable of hybridising to the ligated nucleic acid or primer pairs, for example as described herein, capable of amplifying the ligated nucleic acid in a PCR reaction.

The invention also provides a device that is capable of detecting the relevant chromosome interactions. The device preferably comprises any specific binding agents, probe or primer pair capable of detecting the chromosome interaction, such as any such agent, probe or primer pair described herein.

Publications

The contents of all publications mentioned herein are incorporated by reference into the present specification and may be used to further define the features relevant to the invention.

Specific Embodiments

The EpiSwitch™ platform technology detects epigenetic regulatory signatures of regulatory changes between normal and abnormal conditions at loci. The EpiSwitch™ platform identifies and monitors the fundamental epigenetic level of gene regulation associated with regulatory high order structures of human chromosomes also known as chromosome conformation signatures. Chromosome signatures are a distinct primary step in a cascade of gene deregulation. They are high order biomarkers with a unique set of advantages against biomarker platforms that utilize late epigenetic and gene expression biomarkers, such as DNA methylation and RNA profiling.

10 **EpiSwitch™ Array Assay**

The custom EpiSwitch™ array-screening platforms come in 4 densities of, 15K, 45K, 100K, and 250K unique chromosome conformations, each chimeric fragment is repeated on the arrays 4 times, making the effective densities 60K, 180K, 400K and 1 Million respectively.

15 **Custom Designed EpiSwitch™ Arrays**

The 15K EpiSwitch™ array can screen the whole genome including around 300 loci interrogated with the EpiSwitch™ Biomarker discovery technology. The EpiSwitch™ array is built on the Agilent SurePrint G3 Custom CGH microarray platform; this technology offers 4 densities, 60K, 180K, 400K and 1 Million probes. The density per array is reduced to 15K, 45K, 100K and 250K as each EpiSwitch™ probe is presented as a quadruplicate, thus allowing for statistical evaluation of the reproducibility. The average number of potential EpiSwitch™ markers interrogated per genetic loci is 50; as such the numbers of loci that can be investigated are 300, 900, 2000, and 5000.

EpiSwitch™ Custom Array Pipeline

25 The EpiSwitch™ array is a dual colour system with one set of samples, after EpiSwitch™ library generation, labelled in Cy5 and the other of sample (controls) to be compared/ analyzed labelled in Cy3. The arrays are scanned using the Agilent SureScan Scanner and the resultant features extracted using the Agilent Feature Extraction software. The data is then processed using the EpiSwitch™ array processing scripts in R. The arrays are processed using standard dual colour packages in
30 Bioconductor in R: Limma *. The normalization of the arrays is done using the normalized within Arrays function in Limma * and this done to the on chip Agilent positive controls and EpiSwitch™ positive controls. The data is filtered based on the Agilent Flag calls, the Agilent control probes are removed and the technical replicate probes are averaged, in order for them to be analyzed using Limma *. The probes are modelled based on their difference between the 2 scenarios being

compared and then corrected by using False Discover rate. Probes with Coefficient of Variation (CV) <30% that are <1 or > 1 and pass the $p=0.01$ FDR p-value are used for further screening. To reduce the probe set further Multiple Factor Analysis is performed using the FactorMineR package in R.

- 5 * Note: LIMMA is Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. Limma is a R package for the analysis of gene expression data arising from microarray or RNA-Seq.

10 The pool of probes is initially selected based on adjusted p-value, FC and CV <30% (arbitrary cut off point) parameters for final picking. Further analyses and the final list are drawn based only on the first two parameters (adj p-value; FC).

Publications

15 The contents of all publications mentioned herein are incorporated by reference into the present specification and may be used to further define the features relevant to the invention.

Specific Embodiments

20 The EpiSwitch™ platform technology detects epigenetic regulatory signatures of regulatory changes between normal and abnormal conditions at loci. The EpiSwitch™ platform identifies and monitors the fundamental epigenetic level of gene regulation associated with regulatory high order structures of human chromosomes also known as chromosome conformation signatures. Chromosome signatures are a distinct primary step in a cascade of gene deregulation. They are high order biomarkers with a unique set of advantages against biomarker platforms that utilize late epigenetic and gene expression biomarkers, such as DNA methylation and RNA profiling.

25 *EpiSwitch™ Array Assay*

The custom EpiSwitch™ array-screening platforms come in 4 densities of, 15K, 45K, 100K, and 250K unique chromosome conformations, each chimeric fragment is repeated on the arrays 4 times, making the effective densities 60K, 180K, 400K and 1 Million respectively.

30 *Custom Designed EpiSwitch™ Arrays*

The 15K EpiSwitch™ array can screen the whole genome including around 300 loci interrogated with the EpiSwitch™ Biomarker discovery technology. The EpiSwitch™ array is built on the Agilent SurePrint

G3 Custom CGH microarray platform; this technology offers 4 densities, 60K, 180K, 400K and 1 Million probes. The density per array is reduced to 15K, 45K, 100K and 250K as each EpiSwitch™ probe is presented as a quadruplicate, thus allowing for statistical evaluation of the reproducibility. The average number of potential EpiSwitch™ markers interrogated per genetic loci is 50; as such the numbers of loci that can be investigated are 300, 900, 2000, and 5000.

EpiSwitch™ Custom Array Pipeline

The EpiSwitch™ array is a dual colour system with one set of samples, after EpiSwitch™ library generation, labelled in Cy5 and the other of sample (controls) to be compared/ analyzed labelled in Cy3. The arrays are scanned using the Agilent SureScan Scanner and the resultant features extracted using the Agilent Feature Extraction software. The data is then processed using the EpiSwitch™ array processing scripts in R. The arrays are processed using standard dual colour packages in Bioconductor in R: Limma *. The normalisation of the arrays is done using the normalised within Arrays function in Limma * and this is done to the on chip Agilent positive controls and EpiSwitch™ positive controls. The data is filtered based on the Agilent Flag calls, the Agilent control probes are removed and the technical replicate probes are averaged, in order for them to be analysed using Limma *. The probes are modelled based on their difference between the 2 scenarios being compared and then corrected by using False Discovery Rate. Probes with Coefficient of Variation (CV) $\leq 30\%$ that are ≤ -1.1 or ≥ 1.1 and pass the $p \leq 0.1$ FDR p-value are used for further screening. To reduce the probe set further Multiple Factor Analysis is performed using the FactorMineR package in R.

* Note: LIMMA is Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. Limma is a R package for the analysis of gene expression data arising from microarray or RNA-Seq.

The pool of probes is initially selected based on adjusted p-value, FC and CV $< 30\%$ (arbitrary cut off point) parameters for final picking. Further analyses and the final list are drawn based only on the first two parameters (adj p-value; FC).

Examples

The invention is illustrated by the following non-limiting Examples.

5 Statistical Pipeline

EpiSwitch™ screening arrays are processed using the EpiSwitch™ Analytical Package in R in order to select high value EpiSwitch™ markers for translation on to the EpiSwitch™ PCR platform.

Step 1

- 10 Probes are selected based on their corrected p-value (False Discovery Rate, FDR), which is the product of a modified linear regression model. Probes below p-value ≤ 0.1 are selected and then further reduced by their Epigenetic ratio (ER), probes ER have to be ≤ -1.1 or ≥ 1.1 in order to be selected for further analysis. The last filter is a coefficient of variation (CV), probes have to be below ≤ 0.3 .

Step 2

- 15 The top 40 markers from the statistical lists are selected based on their ER for selection as markers for PCR translation. The top 20 markers with the highest negative ER load and the top 20 markers with the highest positive ER load form the list.

Step 3

- 20 The resultant markers from step 1, the statistically significant probes form the bases of enrichment analysis using hypergeometric enrichment (HE). This analysis enables marker reduction from the significant probe list, and along with the markers from step 2 forms the list of probes translated on to the EpiSwitch™ PCR platform.

The statistical probes are processed by HE to determine which genetic locations have an enrichment of statistically significant probes, indicating which genetic locations are hubs of epigenetic difference.

- 25 The most significant enriched loci based on a corrected p-value are selected for probe list generation. Genetic locations below p-value of 0.3 or 0.2 are selected. The statistical probes mapping to these genetic locations, with the markers from step 2, form the high value markers for EpiSwitch™ PCR translation.

Example 1: A method of determining the chromosome interactions which are relevant to a companion diagnostic that distinguishes between non-responders and responders of methotrexate for the treatment of Rheumatoid Arthritis

5 Source: Glasgow Scottish Educational Research Association (SERA) cohort.

Introduction to and Brief Summary of Example 1

10 Stable epigenetic profiles of individual patients modulate sensitivity of signalling pathways, regulate gene expression, influence the paths of disease development, and can render ineffective the regulatory controls responsible for effective action of the drug and response to treatment. Here we analysed epigenetic profiles of rheumatoid arthritis (RA) patients in order to evaluate its role in defining the non-responders to Methotrexate (MTX) treatment.

15 Reliable clinical prediction of response to first-line disease modifying anti-rheumatic drugs (DMARDs, usually methotrexate (MTX)) in rheumatoid arthritis is not currently possible. Currently the ability to determine response to first line DMARDs (in particular, methotrexate (MTX) is dependent on empiric clinical measures after the therapy.

20 In early rheumatoid arthritis (ERA), it has not been possible to predict response to first line DMARDs (in particular methotrexate (MTX)) and as such treatment decisions rely primarily on clinical algorithms. The capacity to classify drug naïve patients into those that will not respond to first line DMARDs would be an invaluable tool for patient stratification. Here we report that chromosome conformational signatures (highly informative and stable epigenetic modifications that
25 have not previously been described in RA) in blood leukocytes of early RA patients can predict non-responsiveness to MTX treatment.

Methods:

Peripheral blood mononuclear cells (PBMCs) were obtained from DMARD naïve ERA patients recruited
30 in the Scottish early rheumatoid arthritis (SERA) inception cohort. Inclusion in this study was based on diagnosis of RA (fulfilling the 2010 ACR/EULAR Criteria) with moderate to high disease activity (DAS28 ≥ 3.2) and subsequent monotherapy with methotrexate (MTX). DAS28 = Disease Activity Score of 28 joints. EULAR = The European League Against Rheumatism. ACR = American College of Rheumatology. MTX responsiveness was defined at 6 months using the following criteria: Responders – DAS28
35 remission (DAS28 < 2.6) or a good response (DAS28 improvement of > 1.2 and DAS28 ≤ 3.2). Non-

responders - no improvement in DAS28 (≤ 0.6). Initial analysis of chromosome conformational signatures (CCS) in 4 MTX responders, 4 MTX non-responders and 4 healthy controls was undertaken using an EpiSwitch™ array containing 13,322 unique probes covering 309 RA-related genetic loci. Differentiating CCS were defined by LIMMA * linear modeling, subsequent binary filtering and cluster analysis. A validation cohort of 30 MTX responders and 30 non-responders were screened for the differentiating CCS using the EpiSwitch™ PCR platform. The differentiating signature was further refined using binary scores and logistical regression modeling, and the accuracy and robustness of the model determined by ROC ** analysis.

* Note: LIMMA is Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. Limma is a R package for the analysis of gene expression data arising from microarray or RNA-Seq.

** Note: ROC means Receiver Operating Characteristic and refers to ROC curves. An ROC curve is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate against the false positive rate at various threshold settings.

CCS EpiSwitch™ array analysis identified a 30-marker stratifying profile differentiating responder and non-responder ERA patients. Subsequent evaluation of this signature in our validation cohort refined this to a 5-marker CCS signature that was able to discriminate responders and non-responders. Prediction modeling provided a probability score for responders and non-responders, ranging from 0.0098 to 0.99 (0 = responder, 1 = non-responder). There was a true positive rate of 92% (95% confidence interval [95% CI] 75-99%) for responders and a true negative rate of 93% (95% CI 76-99%) for non-responders. Importantly, ROC analysis to validate this stratification model demonstrated that the signature had a predictive power of sensitivity at 92% for NR to MTX.

We have identified a highly informative systemic epigenetic state in the peripheral blood of DMARD naïve ERA patients that has the power to stratify patients at the time of diagnosis. The capacity to differentiate patients *a priori* into non-responders, using a blood-based clinical test, would be an invaluable clinical tool; paving the way towards stratified medicine and justifying more aggressive treatment regimes in ERA clinics.

Detailed Version of Example 1

The capacity to differentiate patients *a priori* into responders (R) and non-responders (NR) would be an invaluable tool for patient stratification leading to earlier introduction of effective treatment. We have used the EpiSwitch™ biomarker discovery platform to identify Chromosome Conformation Signatures (CCS) in blood-derived leukocytes, which are indicative of disease state and MTX responsiveness. Thereby we identified an epigenetic signature contained in the CXCL13, IFNAR1, IL-17A, IL-21R and IL-23 loci that provide the first prognostic molecular signature that enables the stratification of treatment naïve early RA (ERA) patients into MTX R and NR. Importantly, this stratification model had a predictive power of sensitivity at 92% for NR to MTX. This epigenetic RA biomarker signature can distinguish between ERA and healthy controls (HC). This combinatorial, predictive peripheral blood signature can support earlier introduction of more aggressive therapeutics in the clinic, paving the way towards personalized medicine in RA.

RA is a chronic autoimmune disease affecting up to 1% of the global population. Pathogenesis is multifactorial and characterized by primarily immune host gene loci interacting with environmental factors, particularly smoking and other pulmonary stimuli^{1,2,3}. The exposure of a genetically susceptible individual to such environmental factors suggests an epigenetic context for disease onset and progression. Recent studies of chromatin markers (e.g. methylation status of the genome) provide the first evidence of epigenetic differences associated with RA^{4,5,6,7}. However, to date neither genetic associations, nor epigenetic changes, have provided a validated predictive marker for response to a given therapy. Moreover, clinical presentation only weakly predicts the efficacy and toxicity of conventional DMARDs. MTX⁸, the commonest first-choice medication recommended by EULAR (The European League Against Rheumatism) and ACR (American College of Rheumatology) management guidelines, delivers clinically meaningful response rates ranging from 50 to 65% after 6 months of treatment¹¹. Such responses, and especially the rather smaller proportion that exhibits high hurdle responses, cannot currently be predicted in an individual patient. This begets a 'trial and error' based approach to therapeutic regimen choice (mono or combinatorial therapeutics). The ability to predict drug responsiveness in an individual patient would be an invaluable clinical tool, given that response to first-line treatment is the most significant predictor of long-term outcome^{9,10}.

Herein we focused on epigenetic profiling of DMARD-naïve, ERA patients from the Scottish Early Rheumatoid Arthritis (SERA) inception cohort in order to ascertain if there is a stable blood-based epigenetic profile that indicates NR to MTX treatment and thus enables *a priori* identification and

stratification of such patients to an alternate therapeutic. The source Epigenetic modulation can strongly influence cellular activation and transcriptional profiles. Conceivably, the mode of action for a drug could be affected by epigenetically modified loci. We have focused on CCS, also known as long-range chromatin interactions, because they reflect highly informative and stable high-order epigenetic status which have significant implications for transcriptional regulation^{12,13,14}. They also offer significant advantages¹⁵ and early functional links to phenotypic differences¹⁶, and have been reported as informative biomarkers candidates in oncology and other disease areas^{17,18,19}.

We used early RA (ERA) patients provided by the Scottish early rheumatoid arthritis (SERA) inception cohort. Demographic, clinical and immunological factors were obtained at diagnosis and 6 months. Inclusion in this study was based on a diagnosis of RA (fulfilling the 2010 ACR/EULAR Criteria) with moderate to high disease activity ($\text{DAS28} \geq 3.2$) and subsequent monotherapy with MTX. Responders were defined as patients who upon receiving MTX achieved DAS28 remission ($\text{DAS28} < 2.6$) or a good response (DAS28 improvement of > 1.2 and $\text{DAS28} \leq 3.2$) at 6 months. Non-responders were defined as patients who upon receiving MTX had no improvement in DAS28 (≤ 0.6) at 6 months. Blood samples for epigenetic analysis were collected at diagnosis. (DAS28 = Disease Activity Score of 28 joints.)

We used a binary epigenetic biomarker profiling by analysing over 13,322 chromosome conformation signatures (CCS) (13,322 unique probes) across 309 genetic loci functionally linked to RA. CCS, as a highly informative class of epigenetic biomarkers (1), were read, monitored and evaluated on EpiSwitch™ platform which has been already successfully utilized in blood based stratifications of Mayo Clinic cohort with early melanoma (2) and is currently used for predictive stratification of responses to immunotherapies with PD-1/PD-L1.

Identified epigenetic profiles of naïve RA patients were subject to statistical analysis using GraphPad Prism, WEKA and R Statistical language. By using EpiSwitch™ platform and extended cohort of 90 clinical samples we have identified a pool of over 922 epigenetic lead biomarkers, statistically significant for responders, non-responders, RA patients and healthy controls.

To identify a pre-treatment circulating CCS status in ERA patients, 123 genetic loci (Table 1) associated with RA pathogenesis were selected and annotated with chromosome conformations interactions predicted using the EpiSwitch™ *in silico* prediction package²⁰. The EpiSwitch™ *in silico* prediction generated 13,322 high-confidence CCS marker candidates (Table 1). These candidates were used to generate a bespoke discovery EpiSwitch™ array (Figure 5) to screen peripheral blood mononuclear cells isolated at the time of diagnosis (DMARD-naïve) from 4 MTX responders (R) and 4 MTX NR, all

clinically defined after 6 months therapy (Figure 1A, B and Table 2), and 4 healthy controls (HC). To identify the CCS that differentiated R, NR and HC, a LIMMA linear model of the normalized epigenetic load was employed. A total of 922 statistically significant stratifying markers (significance assessed on the basis of adjusted p value and EpiSwitch™ Ratio) were identified. Of the 922 lead markers, 420 were associated with NR, 210 with R and 159 with HC (Fig. 1C). Binary filtering and cluster analysis was applied to the EpiSwitch™ markers to assess the significance of CCS identified. A stepwise hierarchical clustering approach (using Manhattan distance measure with complete linkage agglomeration and taking into account R vs NR, HC vs R & HC vs NR) reduced the number of significant markers from 922 to 65 and finally resulted in a 30-marker stratifying profile (Figure 1D and Table 3).

To refine and validate the CCS signature, the 30 identified markers were screened in a second ERA patient cohort of R and NR (Figure 2A, B and Table 4) in a stepwise approach, using the EpiSwitch™ PCR platform (Figure 5). In the first instance, the entire 30 CCS markers were run in 12 ERA patients (6 R and 6 NR). The best differentiating CCS markers were identified by applying a Chi-squared test for independence with Yate's continuity correction on the binary scores, revealing a 12-marker CCS profile (Table 5). These 12 CCS markers were run on an additional 12 ERA patients (6 R and 6 NR) and the data combined with the previous 12 ERA. Combining the 24 patient samples (12 R and 12 NR) a logistic regression Model in the WEKA classification platform (using 5-fold cross validation to score the discerning power of each marker) was built and run 10 times by random data re-sampling of the initial data set to generate 10 different start points for model generation. The markers with the highest average scores were selected, thus reducing the profile to the 10 best discerning CCS markers (Table 5). The 10 CCS markers were used to probe a further 36 ERA samples (18 R and 18 NR). Combining all data (30 R and 30 NR), and using the same logistical regression and score calculation analysis, revealed a 5 CCS marker signature (IFNAR1, IL-21R, IL-23, IL-17A and CXCL13) that distinguished MTX R from NR (Figure 2C, and Table 5). CCS in the CXCL13 and IL-17A loci were associated with non-responders whilst CCS in the IFNAR1, IL-23 and IL-21R loci were associated with responders. This was an intriguing profile given the central role postulated for the IL-17 axis in human autoimmunity.

Importantly, the composition of the stratifying signature identifies the location of chromosomal conformations that potentially control genetic locations of primary importance for determining MTX response. Principal component analysis (PCA) of the binary scores for the classifying 5 EpiSwitch™ CCS markers provided clear separation of ERA patients based on their MTX response (Figure 2D). The model provided a prediction probability score for responders and non-responders, ranging from 0.0098 to 0.99 (0 = responder, 1 = non-responder). The cut-off values were set at ≤ 0.30 for responders

and ≥ 0.70 for non-responders. The score of ≤ 0.30 had a true positive rate of 92% (95% confidence interval [95% CI] 75-99%) whilst a score of ≥ 0.70 had a true negative response rate of 93% (95% CI 76-99%). The number of observed and predicted patients per response category (R or NR to MTX) is shown in Table 6. With the EpiSwitch™ CCS marker model, 53 patients (88%) were classified as either responder or non-responder.

Table 6. Observed and predicted number of R and NR to MTX monotherapy at 6 months using the EpiSwitch™ CCS model

Observed response	Predicted response		
	Non-responder	Undefined	Responder
Non-responder	25	3	2
Responder	2	4	24

Notes to Table 6: Cut off levels were chosen based on the probability of response to MTX of (approximately) >0.70 for NR and <0.3 for R. NR and R were defined as described in the methods.

In order to test the 'accuracy' and 'robustness of performance' of the logistic classifying model that determined the 5 EpiSwitch™ CSS markers, 150 ROC ** curves (with unique start points) were generated by random data re-sampling of the R and NR data (Figure 3A). This resulted in the data being split into training (66%, equivalent to 6000 known class samples) and test (34%, equivalent to 3000 unknown class samples) groups; importantly the same split is never seen in the data for cross validation. The average discriminative ability (AUC) of the model was 89.9% (95% CI 87-100%), with an average sensitivity (adjusted for response prevalence) for NR of 92% and an average specificity for R of 84%. To determine the predictive capability of the model, the average model accuracy statistics were adjusted for population R/NR to MTX using Bayes prevalence theorem²¹. Using a 55% MTX response rate, the positive predictive value (PPV) was 90.3% whilst the negative predictive value (NPV) was 86.5%. If the response rate was adjusted to 60%, this decreased the PPV to 87% whilst increasing the NPV to 89%.

** Note: ROC means Receiver Operating Characteristic and refers to ROC curves. An ROC curve is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate against the false positive rate at various threshold settings.

As an independent evaluation of the discerning powers of the selected 5 EpiSwitch™ CCS markers, factor analysis of mixed data (FAMD) incorporating 30 HC was performed. This illustrated that the signature not only has the power to differentiate between MTX R and NR but also retains sufficient disease-specific features to differentiate between healthy individuals and RA patients (Figure 3B).

Example 1 - Table 8a - Stratifying between RA-MTX responders and non-responders

Table 8a, and continuation Table 8b, presented hereinafter, show *inter alia* a list of about 54 DNA probes (60mers) and their DNA sequences. These probes represent some of the probes used in Example 1. Without being bound, most of the probes illustrated in Table 8a+8b are thought likely to be significant to / useful in stratifying between RA-MTX responders and RA-MTX non-responders. The shown probes were investigated further by PCR. P Value = Probability value; adj. = adjusted.

Example 1 - Conclusion

In conclusion, our study of the epigenetic profile classification of DMARD naïve ERA patients on the basis of prospective clinical assessment for R/NR has identified a consistent epigenetic signature, which discriminates an epigenetic state that is conducive and non-conductive to MTX response. This is to our knowledge, the first example of a stable and selectively differentiating blood based epigenetic biomarker in early RA patients that appears disease related (versus healthy controls) and that can predict non-responsiveness to first-line MTX therapy. This model offers direct and practical benefits with a validated classifier based on 5 conditional CCS and their detection by the industrial ISO-13485 EpiSwitch™ platform, which has the potential to be routinely available in the near future within clinical practice. Importantly, by adopting this predictive signature it should be possible to stratify MTX naïve ERA patients into R and NR cohorts. This offers the potential to accelerate patient progression through the currently approved treatment strategy for ERA seeking earlier use of effective therapeutics, hence leading to a 'personalised' treatment regime. Furthermore, it is conceivable that alternative CCS signatures are present in RA patients (and patients with other autoimmune diseases) that could be used to justify fast-tracked biological treatment regimes in the clinic. This would have far reaching socio-economic implications, providing more cost effective and robust therapeutic approaches.

*Example 1 - Material and Methods**Example 1 - RA patient population*

ERA patients in this study are part of the Scottish early rheumatoid arthritis (SERA) inception cohort. Demographic, clinical and immunological factors were obtained at diagnosis and 6 months (Table 2).

- 5 Inclusion in the inception cohort was based on clinical diagnosis of undifferentiated polyarthritis or RA (≥ 1 swollen joint) at a secondary care rheumatology unit in Scotland. Exclusion criteria were previous or current DMARD/biological therapy and/or established alternative diagnosis (i.e. psoriatic arthritis, reactive arthritis). Inclusion in this study was based on a diagnosis of RA (fulfilled the 2010 ACR/EULAR criteria for RA) with moderate to high disease activity ($\text{DAS28} \geq 3.2$) and subsequent monotherapy
- 10 with MTX. [DAS28 = Disease Activity Score of 28 joints. EULAR = The European League Against Rheumatism. ACR = American College of Rheumatology.] Responders were defined as patients who upon receiving MTX achieved DAS28 remission ($\text{DAS28} < 2.6$) or a good response (DAS28 improvement of > 1.2 and $\text{DAS28} \leq 3.2$) at 6 months. Non-responders were defined as patients who upon receiving MTX had no improvement in DAS28 (≤ 0.6) at 6 months. Blood samples were collected at diagnosis
- 15 (Baseline) in EDTA tubes and centrifuged to generate a buffy layer containing PBMCs, which was harvested and stored at -80°C . Local ethics committees approved the study protocol and all patients gave informed consent before enrolment into the study.

Example 1 - EpiSwitchTM processing, array and PCR detection. Probe design and locations for EpiSwitchTM assays

- 20 Pattern recognition methodology was used to analyse human genome data in relation to the transcriptional units in the human genome. The proprietary EpiSwitchTM pattern recognition software^{18, 20} provides a probabilistic score that a region is involved in chromatin interaction. Sequences from 123 gene loci were downloaded and processed to generate a list of the 13,322 most
- 25 probable chromosomal interactions. 60mer probes were designed to interrogate these potential interactions and uploaded as a custom array to the Agilent SureDesign website. Sequence-specific oligonucleotides were designed using Primer3²³, at the chosen sites for screening potential markers by nested PCR. Oligonucleotides were tested for specificity using oligonucleotide specific BLAST.

30 *Example 1 - Chromatin Conformation signature analysis from patient PBMC's*

Template preparation: Chromatin from 50 μl of each PBMC sample was extracted using the EpiSwitchTM assay following the manufacturer's instructions (Oxford BioDynamics Ltd). Briefly, the

higher order structures are fixed with formaldehyde, the chromatin extracted, digested with TaqI, dilution and ligation in conditions to maximize intramolecular ligation, and subsequent proteinase K treatment. *EpiSwitch™ microarray*: EpiSwitch™ microarray hybridization was performed using the custom Agilent 8x60k array using the Agilent system, following the manufacturer's instructions (Agilent). Each array contains 55088 probes spots, representing 13,322 potential chromosomal interactions predicted by the EpiSwitch™ pattern recognition software quadruplicated, plus EpiSwitch™ and Agilent controls. Briefly, 1 µg of EpiSwitch™ template was labelled using the Agilent SureTag labelling kit. Processing of labelled DNA was performed. Array analysis was performed immediately after washing using the Agilent scanner and software. In order to compare all the experiments the data was background corrected and normalized. Since each spot in the array is present in quadruplicate, the median of the four spots of each probe in the array was calculated and its log2 transformed value was used for further analysis. The coefficient of variation and *p*-value was calculated for each probe replicate. *EpiSwitch™ PCR detection*: Oligonucleotides were tested on template to confirm that each primer set was working correctly. To accommodate for technical and replicate variations, each sample was processed four times. All the extracts from these four replicates were pooled and the final nested PCR was performed on each sample. This procedure permitted the detection of limited copy-number templates with higher accuracy²⁴. All PCR amplified samples were visualised by electrophoresis in the LabChip® GX from Perkin Elmer, using the LabChip DNA 1K Version2 kit (Perkin Elmer) and internal DNA marker was loaded on the DNA chip according to the manufacturer's protocol using fluorescent dyes. Fluorescence was detected by laser and electropherogram read-outs translated into a simulated band on gel picture using the instrument software. The threshold we set for a band to be deemed positive was 30 fluorescence units and above.

Example 1 - Statistical Methods and Packages

GraphPad Prism and SPSS were used for all statistical analyses of clinical data. The chi-square test and Fisher's exact test (for categorical variables), the t-test for independent samples (for continuous normally distributed variables), and the Mann-Whitney U test (for continuous variables without normal distribution) were used to identify differences. The level of statistical significance was set at 0.05, and all tests were 2-sided. R (and appropriate packages) was used for evaluation of EpiSwitch™ data. This included Stats package for Chi-square test and GLM (logit), ROCR package for ROC curves from WEKA odds probabilities, gplot & stats package in R for Heatmaps. FactorMiner package was used for PCA and Factor plots. Weka was used for Attribute Reduction, data randomisation and re-sampling, Logistic Model Classifier, AUC calculations and model accuracy calculations.

References for Example 1 and for all of the present patent specification

1. Liao, K. P., Alfredsson, L. and Karlson, E. W. Environmental influences on risk for rheumatoid
5 arthritis. *Curr. Opin. Rheumatol.* 21, 279–283 (2009).
2. Bottini, N. & Firestein, G. S. Epigenetics in rheumatoid arthritis: a primer for rheumatologists.
Curr. Rheumatol. Rep. 15, 372 (2013).
3. McInnes, I. B. & Schett, G. The pathogenesis of rheumatoid arthritis. *N. Engl. J. Med.* 365,
2205–19 (2011).
- 10 4. Liu, Y. *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary
of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* 31, 142–7 (2013).
5. Nakano, K., Whitaker, J. W., Boyle, D. L., Wang, W. & Firestein, G. S. DNA methylome signature
in rheumatoid arthritis. *Ann. Rheum. Dis.* 72, 110–17 (2013).
6. De la Rica, L. *et al.* Identification of novel markers in rheumatoid arthritis through integrated
15 analysis of DNA methylation and microRNA expression. *J. Autoimmun.* 41, 6–16 (2013).
7. Viatte, S., Plant, D. & Raychaudhuri, S. Genetics and epigenetics of rheumatoid arthritis. *Nat.*
Rev. Rheumatol. 9, 141–53 (2013).
8. Hider, S. L. *et al.* Can clinical factors at presentation be used to predict outcome of treatment
with methotrexate in patients with early inflammatory polyarthritis? *Ann. Rheum. Dis.* 68, 57–62
20 (2009).
9. Farragher, T. M., Lunt, M., Fu, B., Bunn, D. & Symmons, D. P. M. Early treatment with, and
time receiving, first disease-modifying antirheumatic drug predicts long-term function in patients with
inflammatory polyarthritis. *Ann. Rheum. Dis.* 69, 689–95 (2010).
10. Bakker, M. F. *et al.* Early clinical response to treatment predicts 5-year outcome in RA patients:
25 follow-up results from the CAMERA study. *Ann. Rheum. Dis.* 70, 1099–103 (2011).
11. Barrera, P. *et al.* Drug survival, efficacy and toxicity of monotherapy with a fully human anti-
tumour necrosis factor-alpha antibody compared with methotrexate in long-standing rheumatoid
arthritis. *Rheumatology (Oxford)*. 41, 430–439 (2002).
12. Ling, J. Q. & Hoffman, A. R. Epigenetics of long-range chromatin interactions. *Pediatr. Res.* 61,
30 11R–16R (2007).
13. Deng, W. & Blobel, G. A. Do chromatin loops provide epigenetic gene expression states? *Curr.*
Opin. Genet. Dev. 20, 548–54 (2010).
14. Kadauke, S. & Blobel, G. A. Chromatin loops in gene regulation. *Biochim Biophys Acta.* 1789,
17–25 (2009).

15. Crutchley, J. L., Wang, X. Q. D., Ferraiuolo, M. a & Dostie, J. Chromatin conformation signatures: ideal human disease biomarkers? *Biomark. Med.* 4, 611–29 (2010).
16. Christova, R. *et al.* P-STAT1 mediates higher-order chromatin remodelling of the human MHC in response to IFN γ . *J. Cell Sci.* 120, 3262–3270 (2007).
- 5 17. Watanabe, T. *et al.* Higher-Order Chromatin Regulation and Differential Gene Expression in the Human Tumour Necrosis Factor/Lymphotoxin Locus in Hepatocellular Carcinoma Cells. *Mol. Cell. Biol.* 32, 1529–1541 (2012).
18. Mukhopadhyay, S., Ramadass, A. S., Akoulitchiev, A. & Gordon, S. Formation of distinct chromatin conformation signatures epigenetically regulate macrophage activation. *Int.*
10 *Immunopharmacol.* 18, 7–11 (2013).
19. Harismendy, O. *et al.* 9p21 DNA variants associated with coronary artery disease impair interferon- γ signalling response. *Nature* 470, 264–268 (2011).
20. Bastonini, E. *et al.* Chromatin barcodes as biomarkers for melanoma. *Pigment Cell Melanoma Res.* (2014). doi:10.1111/pcmr.12258.
- 15 21. Rau, R. & Herborn, G. Benefit and risk of methotrexate treatment in rheumatoid arthritis. *Clin. Exp. Rheumatol.* 22, S83–S94 (2004).
22. Kosaka, N. & Ochiya, T. Unraveling the Mystery of Cancer by Secretory microRNA: Horizontal microRNA Transfer between Living Cells. *Front. Genet.* 2, 97 (2011).
23. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist
20 programmers. *Methods Mol Biol.* 132, 365–386 (2000).
24. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* 295, 1306–11 (2002).

Example 1 - Table 1. Selected genes for EpiSwitch™ Array

GENE	Description	Comments	Number of identified EpiSwitch™ sites
ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1	MTX related genes	56
ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2	MTX related genes	84
ADORA2A	Adenosine A2a receptor	MTX related genes	72
AFF3	AF4/FMR2 family, member 3	RA SNP association	140
AMPD1	Adenosine monophosphate deaminase 1	MTX related genes	24
ApoE	Apolipoprotein E	Apolipoproteins	96
ATIC	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase	MTX related genes	32
BLK	B lymphoid tyrosine kinase	RA SNP association	196
BTNL2	Butyrophilin-like 2 (MHC class II associated)	Associated with RA via exome sequencing	44
C5orf30	Chromosome 5 open reading frame 30	RA SNP association	96
CCL2	Chemokine (C-C motif) ligand 2	Cytokines & Chemokines	404
CCL21	Chemokine (C-C motif) ligand 21	Cytokines & Chemokines	28
CCL3	Chemokine (C-C motif) ligand 3	Cytokines & Chemokines	52
CCL5	Chemokine (C-C motif) ligand 5	Cytokines & Chemokines	52
CCR1	Chemokine (C-C motif) receptor 1	Cytokines & Chemokines receptors	172
CCR2	Chemokine (C-C motif) receptor 2	Cytokines & Chemokines receptors	164
CCR6	Chemokine (C-C motif) receptor 6	Cytokines & Chemokines receptors	56
CD28	Cluster of Differentiation 28	RA SNP association	132
CD40	Cluster of Differentiation 40	RA SNP association	148
CD80	Cluster of Differentiation 80	Cell surface	76
CHI3L1	Chitinase 3-like 1 (cartilage glycoprotein-39)	Extracellular	64

Example 1 - Table 1. Selected genes for EpiSwitch™ Array

GENE	Description	Comments	Number of identified EpiSwitch™ sites
CHUK	Conserved helix-loop-helix ubiquitous kinase	NFKB	92
CIITA	Class II, major histocompatibility complex, transactivator	NLR pathway	80
CLEC12A	C-type lectin domain family 12, member A	Other	52
CLEC16A	C-type lectin domain family 16, member A	Other	108
COL2A1	Collagen, type II, alpha 1	Collagens	100
CTLA4	Cytotoxic T-lymphocyte-associated protein 4	RA SNP association	68
CX3CL1	Chemokine (C-X3-C motif) ligand 1	Cytokines & Chemokines	92
CXCL12	Chemokine (C-X-C motif) ligand 12	Cytokines & Chemokines	80
CXCL13	Chemokine (C-X-C motif) ligand 13	Cytokines & Chemokines	80
CXCL8	Chemokine (C-X-C motif) ligand 8	Cytokines & Chemokines	48
CXCR3	Chemokine (C-X-C motif) receptor 3	Cytokines & Chemokines receptors	72
CXCR4	Chemokine (C-X-C motif) receptor 4	Cytokines & Chemokines receptors	56
DHFR	Dihydrofolate reductase	MTX related genes	72
ESR1	Oestrogen receptor 1	FLS MTX responsive genes	140
FCGR2A	Fc fragment of IgG, low affinity IIa, receptor (CD32)	RA SNP association	100
FCGR3B	Fc fragment of IgG, low affinity IIIb, receptor (CD16b)	RA SNP association	192
FCRL3	Fc receptor-like 3	Other	68
FPGS	Folypolyglutamate synthase	MTX related genes	56
HTR2A	5-hydroxytryptamine (serotonin) receptor 2A, G protein-coupled	Other	80
ICAM1	Intercellular adhesion molecule 1	FLS MTX responsive genes	132
ICOS	Inducible T-cell co-stimulator	RA SNP association	200

Example 1 - Table 1. Selected genes for EpiSwitch™ Array

GENE	Description	Comments	Number of identified EpiSwitch™ sites
IFNAR1	Interferon (alpha, beta and omega) receptor 1	Cytokines & Chemokines receptors	80
IFNg	Interferon, gamma	Cytokines & Chemokines	52
IKKB	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta	NFKB	128
IL-10	Interleukin 10	Cytokines & Chemokines	48
IL-15	Interleukin 15	Cytokines & Chemokines	76
IL-17A	Interleukin 17A	Cytokines & Chemokines	32
IL-18	Interleukin 18	Cytokines & Chemokines	64
IL-1a	Interleukin 1 alpha	Cytokines & Chemokines	196
IL-2	Interleukin 2	Cytokines & Chemokines	44
IL-21R	Interleukin 21 receptor	Cytokines & Chemokines receptors	60
IL-23	Interleukin 23	Cytokines & Chemokines	56
IL-23R	Interleukin 23 receptor	Cytokines & Chemokines receptors	104
IL-2RA	Interleukin 2 receptor, alpha	Cytokines & Chemokines receptors	100
IL-2RB	Interleukin 2 receptor, beta	Cytokines & Chemokines receptors	72
IL-32	Interleukin 32	Cytokines & Chemokines	44
IL-4	Interleukin 4	Cytokines & Chemokines	32
IL-4R	Interleukin 4 receptor	Cytokines & Chemokines receptors	76
IL-6	Interleukin 6	Cytokines & Chemokines	48
IL-6ST	Interleukin 6 signal transducer (gp130, oncostatin M receptor)	Cytokines & Chemokines receptors	72

Example 1 - Table 1. Selected genes for EpiSwitch™ Array

GENE	Description	Comments	Number of identified EpiSwitch sites
IL-7	Interleukin 7	Cytokines & Chemokines	72
IL1RN	Interleukin 1 receptor antagonist	MTX related genes	28
IRAK3	Interleukin-1 receptor-associated kinase 3	Signalling	80
IRF5	Interferon regulatory factor 5	Signalling	76
ITGA4	Integrin, alpha 4 (antigen CD49D, alpha 4 subunit of VLA-4 receptor)	Cell surface	100
ITPA	Inosine triphosphatase (nucleoside triphosphate pyrophosphatase)	MTX related genes	56
JAG1	Jagged 1	FLS MTX responsive genes	84
M-CSF	Colony stimulating factor 1	Cytokines & Chemokines	96
MafB	V-maf musculoaponeurotic fibrosarcoma oncogene homolog B	Transcription factors	52
MAL	Mal, T-cell differentiation protein	TLR pathway	68
MEFV	Mediterranean fever	Other	76
MMP14	Matrix metalloproteinase 14	Matrix Metalloproteinases	92
MMP2	Matrix metalloproteinase 2	Matrix Metalloproteinases	212
MMP9	Matrix metalloproteinase 9	Matrix Metalloproteinases	68
MTHFD1	Methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1, methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase	MTX related genes	80
MTHFR	Methylenetetrahydrofolate reductase (NAD(P)H)	MTX related genes	52
MyD88	Myeloid differentiation primary response gene 88	TLR pathway	80
NFAT	Nuclear factor of activated T cells	Transcription factors	204
NFATC2IP	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2 interacting protein	RA SNP association	84
NFKB1	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	NFKB	96
NFKB2	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)	NFKB	64

Example 1 - Table 1. Selected genes for EpiSwitch™ Array

GENE	Description	Comments	Number of identified EpiSwitch™ sites
NFKBIB	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, beta	NFKB	120
NFKBIA	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	NFKB	88
NLRP1	NLR family, pyrin domain containing 1	NLR pathway	108
NLRP3	NLR family, pyrin domain containing 3	NLR pathway	128
PADI4	Peptidyl arginine deiminase, type IV	RA SNP association	168
PRDM1	PR domain containing 1, with ZNF domain	RA SNP association	120
PRKCQ	Protein kinase C, theta	RA SNP association	216
PRKCZ	Protein kinase C, zeta	Other	184
PSTPIP1	Proline-serine-threonine phosphatase interacting protein 1	Cytoskeletal	96
PTGS2	Prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	Signalling	52
PTPN22	Protein tyrosine phosphatase, non-receptor type 22	RA SNP association	196
PXK	PX domain containing serine/threonine kinase	RA SNP association	296
RBPJ	Recombination signal binding protein for immunoglobulin kappa J region	RA SNP association	296
REL	V-rel reticuloendotheliosis viral oncogene homolog A	NFKB	92
RFC-1	Replication factor C (activator 1) 1, 145kDa	MTX related genes	52
RGMB	RGM domain family, member B	FLS MTX responsive genes	80
RUNX1	Runt-related transcription factor 1	RA SNP association	212
SH2B3	SH2B adaptor protein 3	RA SNP association	124
SHMT	Serine hydroxymethyltransferase 1 (soluble)	MTX related genes	68
SLC19A1	Solute carrier family 19 (folate transporter), member 1	MTX related genes	76
SPRED2	Sprouty-related, EVH1 domain containing 2	RA SNP association	336
STAT4	Signal transducer and activator of transcription 4	Signalling	128

Example 1 - Table 1. Selected genes for EpiSwitch™ Array

GENE	Description	Comments	Number of identified EpiSwitch™ sites
SUMO1	SMT3 suppressor of mif two 3 homolog 1	SUMO	132
TAGAP	T-cell activation RhoGTPase activating protein	RA SNP association	92
TLR1	Toll-like receptor 1	TLR pathway	204
TLR2	Toll-like receptor 2	TLR pathway	52
TLR4	Toll-like receptor 4	TLR pathway	52
TNF	Tumour necrosis factor	Cytokines & Chemokines	68
TNFAIP3	Tumour necrosis factor, alpha-induced protein 3	RA SNP association	180
TNFRSF11B	Tumour necrosis factor receptor superfamily, member 11b	Cytokines & Chemokines receptors	80
TNFRSF13C	Tumour necrosis factor receptor superfamily, member 13C	Cytokines & Chemokines receptors	52
TNFRSF14	Tumour necrosis factor receptor superfamily, member 14	RA SNP association	112
TNFRSF17	Tumour necrosis factor receptor superfamily, member 17	Cytokines & Chemokines receptors	44
TNFRSF1A	Tumour necrosis factor receptor superfamily, member 1A	Cytokines & Chemokines receptors	72
TNFRSF1B	Tumour necrosis factor receptor superfamily, member 1B	Cytokines & Chemokines receptors	72
TNFSF11	Tumour necrosis factor (ligand) superfamily, member 11	Cytokines & Chemokines	52
TNFSF13	Tumour necrosis factor (ligand) superfamily, member 13	Cytokines & Chemokines	48
TRAF1	TNF receptor-associated factor 1	RA SNP association	120
TRAF6	TNF receptor-associated factor 6	RA SNP association	72
TYMS	Thymidylate synthetase	MTX related genes	48
WISP3	WNT1 inducible Signalling pathway protein 3	Signalling	88

Example 1 - Table 2. Patient Characteristics – Discovery Cohort

	Baseline		P value	6 months			Healthy control
	Non-responder	Responder		Non-responder	Responder	P value	
Age – years	55±6.1	55±19.7	>0.99	-	-	-	52±13.3
Males – no. (%)	1 (25)	1 (25)	1	-	-	-	3 (38)
Caucasian – no. (%)	4 (100)	4 (100)	-	-	-	-	8 (100)
Body mass index – kg/m ²	29.5±0.96 [§]	25.0±4.88	0.19	-	-	-	-
Patient global assessment (VAS, 0-100 mm)	54.3±33.5	39.3±30.2	0.53	54.5±20.0	9.3±6.2	0.029	-
Physician global assessment (VAS, 0-100 mm)	55±29.7	38.5±17.8	0.38	32.5±20.2	8.8±7.0	0.068	-
Number of swollen joints (0-28)	11.3±5.3	4.8±3.9	0.09	15±10.7	2.0±2.8	0.057	-
Number of tender joints (0-28)	10.5±7.7	4.8±6.4	0.2	11.25±10.6	0.5±1.0	0.029	-
CDAI	32.7±5.2	17.3±9.6	0.03	35.0±21.2	4.3±3.7	0.03	-
DAS28-CRP	5.1±0.2	4.2±0.77	0.06	-	-	-	-
DAS28-ESR	5.5±0.5 [§]	4.6±0.9 [§]	0.4	5.3±1.3	2.8±0.7	0.016	-
RF (IU/ml)	35.4±25.6	321±140 [§]	0.06	-	-	-	-
CCP (U/ml)	10.3±7.2	340±0 [§]	0.06	-	-	-	-
Current smoker – no. (%)	2 (50)	1 (25)	-	-	-	-	-
Previous smoker – no. (%)	1 (25)	1 (25)	-	-	-	-	-
Non-smoker – no. (%)	1 (25)	2 (50)	-	-	-	-	-

The Fisher exact unconditional test is used to assess differences in proportions between the two groups. To examine differences in continuous variables between the two groups, the independent samples t-test or the Mann-Whitney U-test (depending on distribution of data) is used. [§] n=3

Example 1 - Table 3. 65 Selected genes from EpiSwitch™ Array analysis

Gene	Probes*	adjusted p value	EpiSwitch™ ratio	HC_NR_MTX	HC_R_MTX	NR_R_MTX	Association
19_55449062_55	19_55449062_55						
451429_5548496	451429_5548496						
0_55486708_RF	0_55486708_RF	0.079228864	-1.43395525	0	-1	-1	R
C5orf30	C5orf30_Site5_Si te2_FF	0.079228864	-1.24257534	1	-1	-1	R
CHUK	CHUK_Site7_Site 2_RF	0.079228864	-1.32868581	1	-1	-1	R
CXCL13	CXCL13_Site1_Sit e3_RR	0.079228864	-1.29833859	0	-1	-1	R
TLR1	TLR1_Site4_Site7 _FR	0.079228864	-1.43064593	1	-1	-1	R
11_47175706_47	11_47175706_47						
180170_4725150	180170_4725150						
5_47252468_FR	5_47252468_FR	0.083312472	-1.20859706	1	-1	-1	R
C5orf30	C5orf30_Site4_Si te2_FF	0.084204721	-1.20024867	1	-1	-1	R
TLR1	TLR1_Site9_Site2 _FF	0.086622849	-1.37554182	1	-1	-1	R
FCRL3	FCRL3_Site9_Site 7_FF	0.090200643	-1.25121814	1	-1	-1	R
SH2B3	SH2B3_Site6_Sit e5_FF	0.090200643	-1.32868581	1	-1	-1	R

Example 1 - Table 3. 65 Selected genes from EpiSwitch™ Array analysis

Gene	Probes*	adjusted p value	EpiSwitch™ ratio	HC_NR_MTX	HC_R_MTX	NR_R_MTX	Association
12_69705360_69	12_69705360_69						
711928_6979916	711928_6979916						
2_69800678_RF	2_69800678_RF	0.097224197	-1.20580783	1	-1	-1	R
	IL-						
	23R_Site5_Site8_						
IL-23R	FF	0.108787769	-1.26868449	1	-1	-1	R
	CLEC12A_Site6_S						
CLEC12A	ite1_FR	0.112869007	-1.22264028	0	-1	-1	R
	IL-						
	17A_Site3_Site1						
IL-17A	_RR	0.115042065	-1.16473359	0	-1	-1	R
	CXCL8_Site7_Site						
CXCL8	6_FR	0.118123176	-1.13288389	0	-1	-1	R
	MyD88_Site5_Sit						
MyD88	e1_FR	0.129904996	-1.18372449	1	0	-1	R
	PRDM1_Site6_Sit						
PRDM1	e2_RR	0.144057138	-1.19195794	1	-1	-1	R
	MMP2_Site8_Sit						
MMP2	e9_FF	0.146105678	-1.20859706	1	-1	-1	R
	SPRED2_Site4_Si						
SPRED2	te8_RF	0.149371667	-1.38510947	1	-1	-1	R
	C5orf30_Site4_Si						
C5orf30	te8_RF	0.150085134	-1.17826714	1	-1	-1	R

Example 1 - Table 3. 65 Selected genes from EpiSwitch™ Array analysis

Gene	Probes*	adjusted p value	EpiSwitch™ ratio	HC_NR_MTX	HC_R_MTX	NR_R_MTX	Association
19_10294661_10	19_10294661_10						
295285_1037056	295285_1037056						
0_10371551_RR	0_10371551_RR	0.153140631	-1.20859706	1	-1	-1	R
TNFRSF13C	TNFRSF13C_Site3_Site6_FF	0.15333898	-1.20580783	1	-1	-1	R
IL-23	IL-23_Site4_Site5_FF	0.160960834	-1.18099266	0	-1	-1	R
NFKB1B	NFKB1B_Site8_Site9_FF	0.168381727	-1.23114441	1	-1	-1	R
TNFRSF13C	TNFRSF13C_Site1_Site6_FF	0.16921449	-1.1198716	1	-1	-1	R
CD28	CD28_Site5_Site9_RR	0.171723501	-1.14340249	1	-1	-1	R
NFKB1	NFKB1_Site4_Site8_RR	0.185725586	-1.20024867	1	-1	-1	R
CHUK	CHUK_Site3_Site5_RF	0.188137111	-1.13026939	1	-1	-1	R
TLR1	TLR1_Site9_Site3_FF	0.188137111	-1.19747871	1	-1	-1	R
M-CSF	M-CSF_Site5_Site6_FF	0.191292635	-1.20859706	1	-1	-1	R

Example 1 - Table 3. 65 Selected genes from EpiSwitch™ Array analysis

Gene	Probes*	adjusted p value	EpiSwitch™ ratio	HC_NR_MTX	HC_R_MTX	NR_R_MTX	Association
NFKB1B	NFKB1B_Site1_Sit e8_FF	0.1919222112	-1.12766093	1	-1	-1	R
11_47175706_47	11_47175706_47						
180170_4720291	180170_4720291						
0_47204016_FF	0_47204016_FF	0.192002056	-1.20580783	1	-1	-1	R
PRDM1	PRDM1_Site6_Sit e1_RR	0.194604588	-1.18920712	1	-1	-1	R
TNFRSF14	TNFRSF14_Site4_ Site1_RR	0.082014717	1.526259209	0	1	1	NR
SH2B3	SH2B3_Site3_Sit e2_FF	0.0833312472	1.228303149	-1	1	1	NR
MyD88	MyD88_Site2_Sit e4_FR	0.086246871	1.211392737	0	1	1	NR
MafB	MafB_Site2_Site 4_FF	0.090511832	1.170128253	-1	1	1	NR
PRKCZ	PRKCZ_Site6_Sit e3_RF	0.093763087	1.316462719	0	1	1	NR
IFNAR1	IFNAR1_Site2_Sit e4_RR	0.093849223	1.228303149	-1	1	1	NR
NFAT	NFAT_Site2_Site 10_FR	0.093849223	1.208597056	-1	1	1	NR
NFAT	NFAT_Site5_Site 10_RR	0.094393734	1.25411241	-1	1	1	NR

Example 1 - Table 3. 65 Selected genes from EpiSwitch™ Array analysis

Gene	Probes*	adjusted p value	EpiSwitch™ ratio	HC_NR_MTX	HC_R_MTX	NR_R_MTX	Association
MAL	MAL_Site2_Site6_RF	0.095094028	1.274560627	0	1	1	NR
	FCGR2A_Site3_Site6_RR	0.096581892	1.170128253	-1	1	1	NR
IL-32	IL-32_Site5_Site4_RF	0.097224197	1.205807828	0	1	1	NR
	MTHFD1_Site1_Site7_RF	0.114751424	1.175547906	-1	1	1	NR
TLR2	TLR2_Site1_Site5_RR	0.120590183	1.217003514	-1	1	1	NR
	NFAT_Site6_Site10_RR	0.129631525	1.211392737	-1	1	1	NR
ICAM1	ICAM1_Site4_Site9_FR	0.131386096	1.180992661	-1	1	1	NR
	NFAT_Site5_Site10_FR	0.133034069	1.170128253	-1	1	1	NR
MTHFD1	MTHFD1_Site5_Site7_RF	0.144559523	1.156688184	-1	1	1	NR
	MTHFR_Site6_Site4_RR	0.150085134	1.170128253	-1	1	1	NR
ICAM1	ICAM1_Site4_Site1_FF	0.151103565	1.140763716	-1	1	1	NR
	MTHFD1_Site1_Site7_RF	0.114751424	1.175547906	-1	1	1	NR

Example 1 - Table 3. 65 Selected genes from EpiSwitch™ Array analysis

Gene	Probes*	adjusted p value	EpiSwitch™ ratio	HC_NR_MTX	HC_R_MTX	NR_R_MTX	Association
NFAT	NFAT_Site11_Sit e10_RR	0.158903523	1.197478705	-1	1	1	NR
	NFAT_Site10_Sit e9_RF	0.160614052	1.197478705	-1	1	1	NR
MafB	MafB_Site5_Site 2_RF	0.167291268	1.164733586	-1	1	1	NR
	NFAT_Site7_Site 10_RR	0.169766598	1.189207115	-1	1	1	NR
FCGR2A	FCGR2A_Site3_Si te7_RR	0.180386617	1.125058485	-1	1	1	NR
	MafB_Site6_Site 2_RF	0.186948332	1.107008782	-1	1	1	NR
ADORA2A	ADORA2A_Site1_ Site7_FR	0.191209559	1.138131035	-1	1	1	NR
	MMP9_Site2_Sit e3_FR	0.192328613	1.132883885	-1	1	1	NR
COL2A1	COL2A1_Site7_Si te2_FF	0.193661549	1.112136086	-1	1	1	NR
	TNFRSF1B_Site1 _Site7_FR	0.19556991	1.154018752	-1	1	1	NR
FCGR2A	FCGR2A_Site3_Si te2_RR	0.197822331	1.117287138	-1	1	1	NR
	IL- 21R_Site5_Site2_ RR	0.199109911	1.125058485	0	1	1	NR

*Probes were designed based on 3 dimensional orientation of the chromosomal confirmation sites. Hence, these were either FF (Forward-Forward), FR (Forward-Reverse), RF (Reverse-Forward) or RR (Reverse-Reverse).

Key	
HC_NR_MTX	1= loop in HC 0=Not_Relevant "-1"= loop in NR
HC_R_MTX	1= loop in HC 0=Not_Relevant "-1"= loop in R
NR_R_MTX	1= loop in NR 0=Not_Relevant "-1"= loop in R

Example 1 - Table 4. Patient characteristics – Validation Cohort

	Baseline		P value	6 months		P value	Healthy control
	Non-Responder	responder		Non-Responder	responder		
Age – years	58±14.5	54±13.2	0.26	-	-	-	45±15.4
Males – no. (%)	10 (33)	13 (43)	0.6	-	-	-	11 (37)
Caucasian – no. (%)	30 (100)	28 (97) [§]	-	-	-	-	-
Body mass index – kg/m ²	28.3±5.4	27.4±4.6 ^{§§}	0.48	-	-	-	-
Patient global assessment (VAS, 0-100 mm)	48±30.2	62±23.0	0.05	64±23.2	11±12.9	<0.0001	-
[¶] Physician global assessment (VAS, 0-100 mm)	46±22.7	54±21.0	0.19	39±6.4	6.4±6.1	<0.0001	-
Number of swollen joints (0-28)	5.8±3.7	8.3±4.3	0.006	6.0±5.2	0.2±0.48	<0.0001	-
Number of tender joints (0-28)	8.4±6.2	7.9±5.2	0.97	11.6±7.7	0.4±0.72	<0.0001	-
[€] CDAI	23.6±10.9	27.8±9.8	0.13	27.9±12.6	2.3±2.2	<0.0001	-
[#] DAS28-CRP	4.8±1.0	5.1±0.9	0.27	5.0±0.8	1.8±0.44	<0.0001	-
[§] DAS28-ESR	5.2±0.8	5.2±1.0	0.98	5.3±0.8	1.8±0.45	<0.0001	-
[¶] RF (IU/ml)	196±244	138±155	0.48	-	-	-	-
[∞] CCP (U/ml)	244±201	314±798	0.25	-	-	-	-
[#] C-reactive protein (mg/liter)	25.8±33.7	23.4±30.0	0.40	12.7±12.2	5.5±5.6	0.005	-

Example 1 - Table 4. Patient characteristics – Validation Cohort

	Baseline		P value	6 months		P value	Healthy control
	Non-Responder	responder		Non-Responder	responder		
[§] Erythrocyte sedimentation rate (mm/hour)	35±19.8	22.6±16.2	0.02	23±18.6	8.5±5.6	0.0004	-
[¶] Whole Blood cell count	8.4±2.2	7.5±1.7	0.09	7.6±2.4	6.5±1.7	0.07	-
[¶] Lymphocytes	1.9±0.59	1.7±0.78	0.09	1.8±0.76	1.7±0.95	0.31	-
[¶] Monocytes	0.63±0.16	0.59±0.22	0.50	0.59±0.45	0.52±0.13	0.38	-
[¶] Eosinophil	0.18±0.14	0.19±0.13	0.55	0.19±0.15	0.17±0.12	0.89	-
[¶] Platelets	332±107	307±86	0.34	299±103	270±79	0.25	-
Current smoker – no. (%)	10 (33)	4 (14)	-	-	-	-	-
Previous smoker – no. (%)	10 (33)	9 (31)	-	-	-	-	-
Non-smoker – no. (%)	10 (33)	16 (55)	-	-	-	-	-

The Fisher exact unconditional test is used to assess differences in proportions between the two groups. To examine differences in continuous variables

between the two groups, we used the independent samples t-test or the Mann-Whitney U-test (depending on distribution of data). [§] One patient “other”

(non-white, non-South East Asian, non-Indian Sub-Continent, Non-Afro-Caribbean), one patient did not give an answer. ^{§§} n= 25 in responders for BMI, [€]

Baseline - n = 29 non-R, n= 30 R; 6m - n = 30 non-R, n= 29, [¶] Baseline - n = 26 non-R, n= 29 R; 6m - n = 21 non-R, n= 29, [§] Baseline - n = 19 non-R, n= 23 R; 6m

- n = 19 non-R, n= 22, [€] Baseline - n = 13 non-R, n= 23 R, [∞] Baseline - n = 26 non-R, n= 29 R, [¶] Baseline - n = 29 non-R, n= 27 R; 6m - n = 28 non-R, n= 25

Example 1 - Table 5. 12 Selected genes from EpiSwitch™ PCR

Gene	EpiSwitch Marker	adjusted.p.value	EpiSwitch™ ratio	HC_NR_MTX	HC_R_MTX	NR_R_MTX	Association
C5orf30	C5orf30_Site5_Site2_FF	0.079228864	-1.242575344	1	-1	-1	R
IFNAR1	IFNAR1_Site2_Site4_RR	0.093849223	1.228303149	-1	1	1	NR
IL-17A	IL-17A_Site3_Site1_RR	0.115042065	-1.164733586	0	-1	-1	R
CXCL13	CXCL13_Site1_Site3_RR	0.079228864	-1.298338588	0	-1	-1	R
IL-21R	IL-21R_Site5_Site2_RR	0.199109911	1.125058485	0	1	1	NR
IL-23	IL-23_Site4_Site5_FR	0.160960834	-1.180992661	0	-1	-1	R
MafB	MafB_Site6_Site2_RF	0.186948332	1.107008782	-1	1	1	NR
FCGR2A	FCGR2A_Site3_Site2_RR	0.197822331	1.117287138	-1	1	1	NR
CLEC12A	CLEC12A_Site6_Site1_FR	0.112869007	-1.222640278	0	-1	-1	R
PRKCZ	PRKCZ_Site6_Site3_RF	0.093763087	1.316462719	0	1	1	NR
MafB	MafB_Site2_Site4_FF	0.090511832	1.170128253	-1	1	1	NR
C5orf30	C5orf30_Site4_Site2_FF	0.084204721	-1.200248667	1	-1	-1	R

Example 1 - Table 6. Observed and predicted number of R and NR to MTX monotherapy at 6 months using the EpiSwitch™ CCS model.

Observed response	Predicted response		
	Non-responder	Undefined	Responder
Non-responder	25	3	2
Responder	2	4	24

Cut off levels were chosen based on the probability of response to MTX of (approximately) >0.70 for NR and <0.3 for R. NR and R were defined as described in the methods.

Example 1A – RA analysis: MTX responders vs. non-responders, and RA vs. Healthy controls: Work subsequent to Example 1

Following on after Example 1, in Example 1A, a biostatistical hypergeometric analysis was carried out, using the “Statistical Pipeline” method(s) at the beginning of the Examples section in the present specification, to generate further refined DNA probes stratifying between MTX responders vs. MTX non-responders, for RA patients on MTX monotherapy.

Example 1A Results: Table 7 (part a and continuation part b) hereinafter discloses Probe and Loci data for RA-MTX – DNA probes stratifying between responders (R) and non-responders (NR). B = B-statistic (lods or B), which is the log-odds that that gene is differentially expressed. FC is the non-log Fold Change. FC_1 is the non-log Fold Change centred around zero. It is seen that Table 7a includes the sequences of 25 refined preferable DNA probes (60mers) for identifying MTX responders (MTX-R), and of 24 (or 25) refined preferable DNA probes (60mers) for identifying MTX non-responders (MTX-NR), from the hypergeometric analysis. Table 9 (parts a, b and c) hereinafter discloses enriched data from a hypergeometric analysis of RA patients vs. healthy controls (HC), and does not relate to the MTX response in RA patients.

Example 1A - Table 7a. Probe and Loci data for RA-MTX - probes stratifying between responders and non-responders.

FC	FC_1	LS	Loop detected	60 mer
0.5774097	-1.7318725	-1	MTX-R	TGTTTTGGCTGCATAAATGCTCTTTCGAAATAATCATCAAAATATTTTCATTGAC
0.6052669	-1.6521636	-1	MTX-R	CACCCCATCTCCCTTGTGACTCTCTCGATGAATCCATTTTTTTGGAAATAGATGAT
0.6567507	-1.5226477	-1	MTX-R	CACCCCATCTCCCTTGTGACTCTCTCGAACTGTGGCAATTTTAACTTTCAAATTG
0.6624775	-1.5094851	-1	MTX-R	CACCCCATCTCCCTTGTGACTCTCTCGAGGCATGATTTGAGTCTTGACAGAAGTTC
0.6628804	-1.5085678	-1	MTX-R	TGCCAGTATTTATTGAGGATTTTGCATCGAGATTGGTTGCATCATGTTGGCCAGGCT
0.6850588	-1.4597286	-1	MTX-R	TGTTTTTGGCTGCATAAATGCTCTTTCGAACTCATGGGCACAAGCAATCCTCCCACC
0.6868153	-1.4559955	-1	MTX-R	TGCCAGTATTTATTGAGGATTTTGCATCGAACAGATGGAGGGAAGGGGATAGCTCC
0.6890053	-1.4513676	-1	MTX-R	TGCCCTAGAGATCTGTGGAACCTTGAACCTCGAGTCGAGTCAAGAGAGATATCAAGAGCTTCTATCA
0.6943398	-1.4402171	-1	MTX-R	CACCCCATCTCCCTTGTGACTCTCTCGAGGGCAGAATGAGCCTCAGACATCTCCAG
0.6963019	-1.4361587	-1	MTX-R	TCTCTGCCTGATTGCCCTGCCAGAACTTCGATTTGGGCTATAGTGTGTTCCAGTCTAA
0.7008036	-1.4269334	-1	MTX-R	CACCCCATCTCCCTTGTGACTCTCTCGATCTTGAAGAGATCTCTCTTAGCAAAAGC
0.7132593	-1.4020146	-1	MTX-R	CACCCCATCTCCCTTGTGACTCTCTCGAAATATTTTGTGAGCTCCTGCTCAT
0.7141705	-1.4002258	-1	MTX-R	TAGGCGCACATGCACACAGCTCGCCTCTTCGACCCAGGAAGATCCAAAGGGAAGTCTGAG
0.7156204	-1.397389	-1	MTX-R	CCCCACCCCATCCAGGAAATTTGGTTTCGATGAGAGAAAGGCAAGAGAACATGGGGTCT
0.7183721	-1.3920362	-1	MTX-R	TGCCAGTATTTATTGAGGATTTTGCATCGAGTTCGAGTTCGAGATCTCTAGGGCA
0.7189408	-1.390935	-1	MTX-R	CTAAAAATTACATCCAGGAAATGAGATATCGAAAGAAAGACATTTATGCGCCCAAAAAACA
0.722487	-1.384108	-1	MTX-R	TAGGCGCACATGCACACAGCTCGCCTCTTCGATGTACAAGCTGCCTATTGATAGACTTTC
0.7254458	-1.3784627	-1	MTX-R	AAAGTTGTGCAATCAGGCAAGTCAAGATTCGAAAGAAAGACATTTATGCGAGCCAAAAACA
0.7374119	-1.3560941	-1	MTX-R	CACCCCATCTCCCTTGTGACTCTCTCGAGTGGTGGAGCAGCCAAACCCAGGGTTCACT
0.7374768	-1.3559748	-1	MTX-R	GGGTCTTGCTATGTTGCCAGGCTGGCCTCGAGATCAGCCTGGGCAACACCGGTGAAAAACC
0.738555	-1.3539954	-1	MTX-R	CTGGTTAGTCTTGGGAGAGTGTATGTGTGAGTTAAGCCATCTGCAATAGCAAGAGAG
0.7415639	-1.3485014	-1	MTX-R	AGCCTTGATCCAGGGATGAAGCCCACTCGAGATATAGATTGAGCCCCAGTTTTTGAG
0.7422652	-1.3472274	-1	MTX-R	ATCGTGTGGGCTGTGTGTGGCAGACTGTTCGAAATCGGAAGCCTCTCTGAAGGTCCAAGG

Example 1A - Table 7a. Probe and Loci data for RA-MTX - probes stratifying between responders and non-responders.

FC	FC_1	LS	Loop detected	60 mer
0.7430431	-1.3458169	-1	MTX-R	TGCCAGTATTTATTGAGGATTTTTCATCGAATCTCTGGTTTATATCCCAATCATTTGT
0.7432273	-1.3454835	-1	MTX-R	CACCCCATCTCCCTTTGCTGACTCTCTCGATATTTGGTGATATTTCAAAGGGTACTTGA
1.6553355	1.65533547	1	MTX-NR	TGATCACTGTTTCTATGAGGATACAGCTCGAGGGGCGAGGGGGGCTCTGGGCCAGGGCG
1.4321012	1.43210121	1	MTX-NR	AACCTATGATTTCTAATCTTGAATGCTGTCGATCTATGAGGAAATGCCCCAGCCTCCCA
1.4179763	1.41797626	1	MTX-NR	CATAATGCATGTGCATGAAAACCTAATCTTCGATCTATGAGGAAATGCCCCAGCCTCCCA
1.4150017	1.41500165	1	MTX-NR	ATCAGTAAAGCTGGTCAGCTACCCATGAATCGATCTATGAGGAAATGCCCCAGCCTCCCA
1.3755396	1.37553964	1	MTX-NR	GTGTCCCAATTTCTAGTCACTGTGAATCGACCTCGCGGGGAGGGGTGCCAGGCCGCATC
1.366009	1.36600904	1	MTX-NR	CCGGGGCTTCCTGTTAAGAATCTTTGTCGATCTATGAGGAAATGCCCCAGCCTCCCA
1.3611955	1.36119553	1	MTX-NR	GTCTTGAAGAAGGACTAATGCTTAGTATCGAGTGCAGCGCGGGTGGGCCAGCACTGCTG
1.3408009	1.34080092	1	MTX-NR	GTTCAATTTAAACATTTTATTATGTATATTCGAGGGGCCAGGCTTTTATACCCCATCTGA
1.3350815	1.33508153	1	MTX-NR	TTCTCCACAGCCGGCGTCTTGGCAGTCGAGGGGCGAGGGGGCGGCTCTGGGGCCAGGGCG
1.3191431	1.31914307	1	MTX-NR	GCAACACATACAACGACTAATCTCTTTTCGACGCCGAGGAGCTCTGCAGTGGGGGCGTA
1.3183444	1.31834441	1	MTX-NR	GTAGGTGCTGAGTAAGTGAGCACTTGCTCGAGGGGCGAGGGGGCGGTCTGGGCCAGGGCG
1.3164851	1.31648512	1	MTX-NR	CAGAAAGACCTTGCAATCATACGGTGCTTCGACGCCGAGGAGCTCTGCAGTGGGGGCGTA
1.3056925	1.3056925	1	MTX-NR	TACTGTGCTGTGCTCGTCAAAAGAGTATGTCGATCTATGAGGAAATGCCCCAGCCTCCCA
1.2876529	1.2876529	1	MTX-NR	CAGAAATTAATCAAAATGCAAGTGACCCCTCGACCCACCCCAAGGGCTGAGGAGTGGGGCAC
1.2777853	1.27778527	1	MTX-NR	AAGGGACCTAGTCCCTATTAAAGATTCTCGAGGGGGCCAGGCTTTTATACCCCATCTGA
1.2773474	1.2773474	1	MTX-NR	CCTGCCGAGACACGGGACGTGGGATTGCTCGATCTATGAGGAAATGCCCCAGCCTCCCA
1.2754233	1.2754233	1	MTX-NR	CCAAAGCTCGCTTTCTTAACCACATACTGTCGAGGGGGCCAGGCTTTTATACCCCATCTGA
1.2747737	1.27477371	1	MTX-NR	TGAATTGTGAGCGTAAGAAATTTATATCTCGAAGTTTGTGAACCTGGCAGGTGGACGGGA
1.2710171	1.2710171	1	MTX-NR	ACCTGATCTGGGGAAGATTAGGAATTGTCGAAACCAATTCCTGGGATGGGGGTGGGGG

Example 1A - Table 7a. Probe and Loci data for RA-MTX - probes stratifying between responders and non-responders.

FC	FC_1	LS	Loop detected	60 mer
1.2689263	1.26892631	1	MTX-NR	GCAAGAGGATCTCTTGAGGCCAGGAGTTCGAGGGGCCAGGCTTTTATACCCCATCTGA
1.2665372	1.2665372	1	MTX-NR	TATCAAGTGATCCAAAAGGCTGCCAGTGTGAGGGGCGAGGGGGCTCTGGGCCAGGG
1.2648953	1.26489531	1	MTX-NR	AAGGGACCTAGTCCCTATTAAGATTTCGAAACCAATTTCTGGGATGGGGGTGGGGG
1.2592485	1.25924848	1	MTX-NR	TATGGACTTTGTAGTCTCATATCAAAGCTCGAAACCAATTTCTGGGATGGGGGTGGGGG
1.2559537	1.25595366	1	MTX-NR	AAAAATAATCTGGCTCTACACTTAGGATTGGAACCAATTTCTGGGATGGGGGTGGGGG

Example 1A - Table 7b. Probe And Loci data for RA-MTX

		Probe Location					4 kb Sequence Location				
FC	FC_1	Chr	Start1	End1	Start2	End2	Chr	Start1	End1	Start2	End2
0.577409	7	12	69702274	69702303	69759619	69759648	12	69702274	69706273	69759619	69763618
0.605266	9	7	22743265	22743294	22801876	22801905	7	22739295	22743294	22797906	22801905
0.656750	7	7	22743265	22743294	22769055	22769084	7	22739295	22743294	22769055	22773054
0.662477	5	7	22743265	22743294	22757576	22757605	7	22739295	22743294	22757576	22761575
0.662880	4	1	67644699	67644728	67729398	67729427	1	67640729	67644728	67725428	67729427
0.685058	8	12	69702274	69702303	69805129	69805158	12	69702274	69706273	69805129	69809128
0.686815	3	1	67644699	67644728	67672222	67672251	1	67640729	67644728	67672222	67676221
0.689005	3	1	67673763	67673792	67752422	67752451	1	67669793	67673792	67748452	67752451
0.694339	8	7	22743265	22743294	22766800	22766829	7	22739295	22743294	22762830	22766829
0.696301	9	4	123383001	123383030	123399247	123399276	4	123379031	123383030	123399247	123403246
0.700803	6	7	22743265	22743294	22765456	22765485	7	22739295	22743294	22765456	22769455
0.713259	3	7	22718635	22718664	22743265	22743294	7	22718635	22722634	22739295	22743294
0.714170	5	12	48397660	48397689	48423816	48423845	12	48397660	48401659	48423816	48427815

Example 1A - Table 7b. Probe And Loci data for RA-MTX

FC	FC_1	Chr	Probe location				Chr	4 kb Sequence location			
			Start1	End1	Start2	End2		Start1	End1	Start2	End2
0.715620	4	17	32738857	32738886	32777305	32777334	17	32738857	32742856	32777305	32781304
0.718372	1	1	67644699	67644728	67673763	67673792	1	67640729	67644728	67669793	67673792
0.718940	8	12	69702274	69702303	69766052	69766081	12	69702274	69706273	69762082	69766081
0.722487	1	12	48397660	48397689	48412400	48412429	12	48397660	48401659	48412400	48416399
0.725445	8	12	69702274	69702303	69806507	69806536	12	69702274	69706273	69802537	69806536
0.737411	9	7	22743265	22743294	22773903	22773932	7	22739295	22743294	22769933	22773932
0.737476	8	19	55449063	55449092	55486679	55486708	19	55449063	55453062	55482709	55486708
0.738555	1	17	32622187	32622216	32745745	32745774	17	32618217	32622216	32745745	32749744
0.741563	9	13	43129388	43129417	43181041	43181070	13	43125418	43129417	43181041	43185040
0.743043	1	10	104130466	104130495	104156468	104156497	10	104126496	104130495	104152498	104156497
0.743227	3	1	67614064	67614093	67644699	67644728	1	67614064	67618063	67640729	67644728
1.655335	5	7	22743265	22743294	22798802	22798831	7	22739295	22743294	22798802	22802801
1.432101	2	1	2460436	2460465	2486982	2487011	1	2456466	2460465	2486982	2490981
1.417976	3	10	6391740	6391769	6577853	6577882	10	6391740	6395739	6577853	6581852
		10	6520005	6520034	6577853	6577882	10	6516035	6520034	6577853	6581852

Example 1A - Table 7b. Probe And Loci data for RA-MTX

		Probe location						4 kb Sequence location					
FC	FC_1	Chr	Start1	End1	Start2	End2	Chr	Start1	End1	Start2	End2		
1.4150017	1.41500165	10	6427823	6427852	6577853	6577882	10	6427823	6431822	6577853	6581852		
1.3755396	1.37553964	18	74845065	74845094	74866978	74867007	18	74845065	74849064	74863008	74867007		
1.366009	1.36600904	10	6470268	6470297	6577853	6577882	10	6466298	6470297	6577853	6581852		
1.3611955	1.36119553	20	44704386	44704415	44720665	44720694	20	44700416	44704415	44716695	44720694		
1.3408009	1.34080092	17	32551069	32551098	32617664	32617693	17	32551069	32555068	32617664	32621663		
1.3350815	1.33508153	1	2486982	2487011	2540813	2540842	1	2486982	2490981	2536843	2540842		
1.3191431	1.31914307	12	66647072	66647101	66696510	66696539	12	66647072	66651071	66696510	66700509		
1.3183444	1.31834441	1	2476023	2476052	2486982	2487011	1	2472053	2476052	2486982	2490981		
1.3164851	1.31648512	12	66663907	66663936	66696510	66696539	12	66663907	66667906	66696510	66700509		
1.3056925	1.30569251	10	6556987	6557016	6577853	6577882	10	6556987	6560986	6577853	6581852		
1.2876529	1.28765291	12	6268999	6269028	6304632	6304661	12	6268999	6272998	6300662	6304661		
1.2777853	1.27778527	17	32617664	32617693	32708031	32708060	17	32617664	32621663	32704061	32708060		
1.2773474	1.27734741	10	6442502	6442531	6577853	6577882	10	6442502	6446501	6577853	6581852		

Example 1A - Table 7b. Probe And Loci data for RA-MTX

		Probe location						4 kb Sequence location					
FC	FC_1	Chr	Start1	End1	Start2	End2	Chr	Start1	End1	Start2	End2		
1.275423	3	17	32529051	32529080	32617664	32617693	17	32525081	32529080	32617664	32621663		
1.274773	7	19	45364170	45364199	45397229	45397258	19	45360200	45364199	45397229	45401228		
1.271017	1	17	32689356	32689385	32738857	32738886	17	32685386	32689385	32738857	32742856		
1.266537	2	1	2486982	2487011	2556784	2556813	1	2486982	2490981	2552814	2556813		
1.264895	3	17	32708031	32708060	32738857	32738886	17	32704061	32708060	32738857	32742856		
1.259338	2	1	11042009	110420126	11047238	11047241	1	11041612	11042012	11047238	11047638		
1.259248	5	17	32553720	32553749	32738857	32738886	17	32549750	32553749	32738857	32742856		
1.255953	7	17	32522613	32522642	32738857	32738886	17	32522613	32526612	32738857	32742856		

Table 7c – continuation of Tables 7a and 7b

probe	Gene Locus	Pro be_ Cou nt_ Tot al	Pro be_ Cou nt_ Sig	Hyper G_ Stats	FDR_ Hyper G	Pe r- ce nt _ Sig	logFC	Ave Expr	t	P. Value	Adj. P. Value	B	FC	FC_1	L S	Loop de te cted
12_69702273_69705360_69759618_69766081_RR	12_69702273_69705360_69759618_69766081	4	2	0.0345 76041	0.5186 40615	50	0.7923 32744	0.7923 32744	6.3527 96842	0.0015 40038	0.2362 361	0.5257 34091	0.5774 09703	1.7318 72526	- 1	MTX -R
IL-6_Site4_Site5_FF	IL-6	48	13	7.18E-05	0.0145 30844	27. 08	0.7243 56533	0.7243 56533	4.7071 12783	0.0055 90201	0.2490 35946	1.6522 57403	0.6052 66944	1.6521 63579	- 1	MTX -R
IL-6_Site4_Site2_FF	IL-6	48	13	7.18E-05	0.0145 30844	27. 08	0.6065 82168	0.6065 82168	6.4603 94591	0.0014 29141	0.2362 361	0.4648 21575	0.6567 50743	1.5226 47688	- 1	MTX -R
IL-6_Site4_Site3_FF	IL-6	48	13	7.18E-05	0.0145 30844	27. 08	0.5940 56548	0.5940 56548	8.5836 74236	0.0003 91843	0.2362 361	0.4977 76542	0.6624 77542	1.5094 85133	- 1	MTX -R
IL-23R_Site4_Site2_FF	IL-23R	104	19	0.0005 50011	0.0548 90393	18. 27	0.5931 79555	0.5931 79555	4.1115 39379	0.0096 61387	0.2554 84712	2.1656 8129	0.6628 80374	1.5085 67818	- 1	MTX -R
12_69702273_69705360_69805128_69806536_RR	12_69702273_69705360_69805128_69806536	4	2	0.0345 76041	0.5186 40615	50	0.5457 00188	0.5457 00188	11.326 82228	0.0001 06595	0.2362 361	1.2726 74673	0.6850 5884	1.4597 28628	- 1	MTX -R
IL-23R_Site4_Site3_FF	IL-23R	104	19	0.0005 50011	0.0548 90393	18. 27	0.5420 05944	0.5420 05944	5.4286 9826	0.0030 62642	0.2382 48996	1.1098 64705	0.6868 15287	1.4559 95548	- 1	MTX -R
IL-23R_Site3_Site7_FF	IL-23R	104	19	0.0005 50011	0.0548 90393	18. 27	0.5374 12982	0.5374 12982	5.1142 55946	0.0039 5047	0.2456 48426	1.3361 15162	0.6890 05315	1.4513 67613	- 1	MTX -R

IL-6_Site4_Site1_FF	IL-6	48	13	7.18E-05	0.0145	27.08	0.5262	0.5262	0.5262	9.1863	0.0002	0.2362	0.7041	0.6943	1.4402	-	MTX -R
IL-2_Site2_Site4_FF	IL-2	44	7	0.0591	0.7726	15.91	0.5222	0.5222	0.5222	5.7183	0.0024	0.2362	0.9143	0.6963	1.4361	-	MTX -R
IL-6_Site4_Site1_FF	IL-6	48	13	7.18E-05	0.0145	27.08	0.5129	0.5129	0.5129	7.3650	0.0007	0.2362	0.0032	0.7008	1.4269	-	MTX -R
IL-6_Site6_Site4_RF	IL-6	48	13	7.18E-05	0.0145	27.08	0.4875	0.4875	0.4875	10.391	0.0001	0.2362	1.0516	0.7132	1.4020	-	MTX -R
COL2A1_Site2_Site5_RR	COL2A1	100	15	0.0132	0.4884	15.36	0.4856	0.4856	0.4856	5.3786	0.0031	0.2382	1.1448	0.7141	1.4002	-	MTX -R
COL2_Site6_Site14_RR	COL2	404	58	9.15E-06	0.0037	36.36	0.4827	0.4827	0.4827	8.4676	0.0004	0.2362	0.4551	0.7156	1.3973	-	MTX -R
IL-23R_Site4_Site3_FF	IL-23R	104	19	0.0005	0.0548	18.27	0.4771	0.4771	0.4771	4.6788	0.0057	0.2490	1.6752	0.7183	1.3920	-	MTX -R
12_69702273_69705360_69759618_69766081_RF	12_69702273_69705360_69759618_69766081	4	2	0.0345	0.5186	50.40	0.4760	0.4760	0.4760	6.9331	0.0010	0.2362	0.2128	0.7189	1.3909	-	MTX -R
COL2A1_Site2_Site4_RR	COL2A1	100	15	0.0132	0.4884	15.36	0.4689	0.4689	0.4689	4.9698	0.0044	0.2473	1.4451	0.7224	1.3841	-	MTX -R
12_69702273_69705360_69805128_69806536_RF	12_69702273_69705360_69805128_69806536	4	2	0.0345	0.5186	50.40	0.4630	0.4630	0.4630	8.2641	0.0004	0.2362	0.3780	0.7254	1.3784	-	MTX -R
IL-6_Site4_Site2_FF	IL-6	48	13	7.18E-05	0.0145	27.08	0.4394	0.4394	0.4394	9.2966	0.0002	0.2362	0.7393	0.7374	1.3560	-	MTX -R

19_55449062_55451429_55484960_55486708_RF	19_55449062_55451429_55484960_55486708	4	2	0.0345	0.5186	0.4393	0.4393	0.4393	0.4393	3.3433	0.0211	0.2949	2.9239	0.7374	1.3559	-	MTX
					40615	50				80062	28841	434	26031	76825	74814	1	-R
CCL2_Site10_Site13_FR	CCL2	404	58	9.15E-06	0.0037	14.	0.4372	0.4372	0.4372	6.9610	0.0010	0.2362	0.1987	0.7385	1.3539	-	MTX
					05017	36	22819	22819	22819	47822	22576	361	30934	54956	95383	1	-R
TNFSF11_Site4_Site2_FR	TNFSF11	52	12	0.0006	0.0548	23.	0.4313	0.4313	0.4313	3.6909	0.0146	0.2777	2.5671	0.7415	1.3485	-	MTX
					90393	08	57024	57024	57024	11039	6314	2544	90834	63929	01404	1	-R
NFKB2_Site5_Site2_FF	NFKB2	54	9	0.0266	0.5186	16.	0.4299	0.4299	0.4299	7.2809	0.0008	0.2362	0.0426	0.7422	1.3472	-	MTX
					40615	67	9336	9336	9336	58467	34343	361	2056	65202	27376	1	-R
IL-23R_Site5_Site4_RF	IL-23R	104	19	0.0005	0.0548	18.	0.4284	0.4284	0.4284	5.6230	0.0026	0.2362	0.9773	0.7430	1.3458	-	MTX
					90393	27	82185	82185	82185	09709	31353	361	92524	43107	16939	1	-R
IL-6_Site4_Site5_FR	IL-6	48	13	7.18E-05	0.0145	27.	0.4281	0.4281	0.4281	7.9572	0.0005	0.2362	0.2555	0.7432	1.3454	-	MTX
					30844	08	24668	24668	24668	32876	55975	361	68458	27265	83471	1	-R
TNFRSF14_Site4_Site1_FR	TNFRSF14	112	14	0.0638	0.7840	12.	0.7271	0.7271	0.7271	3.4991	0.0178	0.2866	2.7611	1.6553	1.6553	-	MTX
					61767	5	23624	23624	23624	9083	94673	24284	97677	35471	35471	1	-NR
PRKCQ_Site11_Site4_RR	PRKCQ	213	31	0.0008	0.0575	14.	0.5181	0.5181	0.5181	3.4418	0.0190	0.2891	2.8206	1.4321	1.4321	-	MTX
					76386	55	33451	33451	33451	02618	15331	91715	09109	01206	01206	1	-NR
PRKCQ_Site7_Site4_FR	PRKCQ	213	31	0.0008	0.0575	14.	0.5038	0.5038	0.5038	3.5630	0.0167	0.2829	2.6958	1.4179	1.4179	-	MTX
					76386	55	33375	33375	33375	03996	36154	50401	57596	76256	76256	1	-NR
PRKCQ_Site9_Site4_RR	PRKCQ	213	31	0.0008	0.0575	14.	0.5008	0.5008	0.5008	3.9015	0.0118	0.2663	2.3620	1.4150	1.4150	-	MTX
					76386	55	0374	0374	0374	43743	59009	7802	04516	01654	01654	1	-NR
18_74845064_74846657_74864995_74867007_RF	18_74845064_74846657_74864995_74867007	4	2	0.0345	0.5186	50	0.4599	0.4599	0.4599	3.6256	0.0156	0.2829	2.6324	1.3755	1.3755	-	MTX
					40615	50	97712	97712	97712	2346	82006	50401	82122	39636	39636	1	-NR

PRKCQ_Site2_Site4_FR	PRKCQ	213		31	0.0008	0.0575	14.	0.4499	0.4499	0.4499	3.4940	0.0179	0.2866	-	1.3660	1.3660	1.3660	MTX -NR
CD40_Site10_Site9_FF	CD40	142		17	0.0622	0.7840	11.	0.4448	0.4448	0.4448	3.5963	0.0161	0.2829	-	1.3611	1.3611	1.3611	MTX -NR
CCL2_Site11_Site10_RR	CCL2	404		58	9.15E-06	0.0037	14.	0.4230	0.4230	0.4230	4.0374	0.0103	0.2564	-	1.3408	1.3408	1.3408	MTX -NR
TNFRSF14_Site1_Site8_RF	TNFRSF14	112		14	0.0638	0.7840	12.	0.4169	0.4169	0.4169	3.3953	0.0199	0.2894	-	1.3350	1.3350	1.3350	MTX -NR
IRAK3_Site2_Site5_RR	IRAK3	75		11	0.0360	0.5216	14.	0.3996	0.3996	0.3996	4.7783	0.0052	0.2490	-	1.3191	1.3191	1.3191	MTX -NR
TNFRSF14_Site6_Site1_FR	TNFRSF14	112		14	0.0638	0.7840	12.	0.3987	0.3987	0.3987	3.5466	0.0170	0.2839	-	1.3183	1.3183	1.3183	MTX -NR
IRAK3_Site4_Site5_RR	IRAK3	75		11	0.0360	0.5216	14.	0.3966	0.3966	0.3966	6.1294	0.0018	0.2362	-	1.3164	1.3164	1.3164	MTX -NR
PRKCQ_Site3_Site4_RR	PRKCQ	213		31	0.0008	0.0575	14.	0.3848	0.3848	0.3848	4.1304	0.0094	0.2554	-	1.3056	1.3056	1.3056	MTX -NR
12_6268998_6272753_6301795_6304661_RF	12_6268998_6272753_6301795_6304661	2		2	0.0064	0.2892	10	0.3647	0.3647	0.3647	3.5905	0.0162	0.2829	-	1.2876	1.2876	1.2876	MTX -NR
CCL2_Site10_Site5_RF	CCL2	404		58	9.15E-06	0.0037	14.	0.3536	0.3536	0.3536	4.3788	0.0075	0.2554	-	1.2777	1.2777	1.2777	MTX -NR
PRKCQ_Site8_Site4_RR	PRKCQ	213		31	0.0008	0.0575	14.	0.3531	0.3531	0.3531	4.9818	0.0044	0.2473	-	1.2773	1.2773	1.2773	MTX -NR

CCL2_Site12_Site10_FR	CCL2		404	58	9.15E-06	0.0037	14.	0.3509	0.3509	0.3509	4.5280	0.0065	0.2510	1.8000	1.2754	1.2754	23299	1	MTX-NR
ApoE_Site3_Site6_FR	ApoE		96	17	0.0015	0.0816	17.	0.3502	0.3502	0.3502	5.5579	0.0027	0.2362	1.0211	1.2747	1.2747	7371	1	MTX-NR
CCL2_Site7_Site6_FR	CCL2		404	58	9.15E-06	0.0037	14.	0.3459	0.3459	0.3459	3.5563	0.0168	0.2836	2.7026	1.2710	1.2710	17097	1	MTX-NR
CCL2_Site2_Site10_FR	CCL2		404	58	9.15E-06	0.0037	14.	0.3436	0.3436	0.3436	4.8095	0.0051	0.2490	1.5701	1.2689	1.2689	26312	1	MTX-NR
TNFRSF14_Site1_Site9_RF	TNFRSF14		112	14	0.0638	0.7840	12.	0.3408	0.3408	0.3408	3.7341	0.0140	0.2766	2.5244	1.2665	1.2665	37198	1	MTX-NR
CCL2_Site5_Site6_FR	CCL2		404	58	9.15E-06	0.0037	14.	0.3390	0.3390	0.3390	4.1920	0.0089	0.2554	2.0925	1.2648	1.2648	95314	1	MTX-NR
M-CSF_Site8_Site3_FR	M-CSF		96	13	0.0426	0.5951	13.	0.3326	0.3326	0.3326	4.6055	0.0061	0.2490	1.7354	1.2593	1.2593	38177	1	MTX-NR
CCL2_Site11_Site6_FR	CCL2		404	58	9.15E-06	0.0037	14.	0.3325	0.3325	0.3325	3.9356	0.0114	0.2629	2.3295	1.2592	1.2592	48484	1	MTX-NR
CCL2_Site12_Site6_FR	CCL2		404	58	9.15E-06	0.0037	14.	0.3287	0.3287	0.3287	3.8761	0.0121	0.2672	2.3862	1.2559	1.2559	53655	1	MTX-NR

Example 1 - Table 8a - Stratifying between RA-MTX responders and non-responders

Probes				Probe sequence	
Probes	NR_R_P.Value	NR_R_adj.P.Val	60 mer		
TNFRSF14_Site4_Site1_FR	0.001232118	0.079419805	TGATCACTGTTTCTTAIGAGGATACAGCTCGAGGGGCAGGGGGCGGTCTCTGGGCCAGGCG	WO 2016/207647	
TNFRSF14_Site4_Site1_RR	0.002061691	0.082014717	AACCTGGAGAACGCCAAGCGCTTCGCCATCGAGGGGCAGGGGGCGGTCTCTGGGCCAGGCG		
TNFRSF1A_Site2_Site5_FR	0.004469941	0.093849223	CTACCTTTGTGGCACTTGGTACAGCAAATCGACGGGCCCCCGTGAGGCGGGGGCGGGACCC		
TNFRSF1A_Site1_Site5_FR	0.005468033	0.09532964	CATCAATTATAACTCACCTTACAGATCATCGACGGGCCCGTGAGGCGGGGGCGGGACCC		
TNFRSF14_Site4_Site8_FR	0.005244102	0.094393734	TGATCACTGTTTCTTAIGAGGATACAGCTCGAAGATTAGGTAAAGGTGGGGACGCGGAGA		
RUNX1_Site7_Site2_RR	0.001313112	0.079419805	GAAAGGTAATTGCCCCCAATAATTATTTTCGAAACAGATCGGGGGCTCGGGTTACACAC		
TNFRSF14_Site1_Site8_RF	0.003725772	0.090200643	TTCTCCACAGCCGGCGTCTTGGCAGTCGAGGGGCAGGGGGCGGTCTCTGGGCCAGGCG		
18_74845064_74846657_74864995_74867007_RF	0.001604249	0.079419805	CGTGTCCCAATTTCTAGTGCACTGTGAATCGACCTCGCGGGAGGGGTGCCAGGCCGCAT		
PRKCZ_Site8_Site6_FR	1.26726E-05	0.079228864	CCCTCTCTCTAAAAGGTCTCAACATCACTCGACTGGAGAGCCCCGGGCTCGCGCCGCTT		
RUNX1_Site5_Site2_RR	0.000540863	0.079228864	GTTTCCCTTGATGCTCAGAGAAAGGCCCTCGAAACAGATCGGGGGCTCGGGTTACACAC		
PRKCQ_Site7_Site4_FR	0.003958472	0.090816122	CATAATGCAATGTCATGAAACTAATCTTCGATCTATGAGGAAATGCCCCCAGCCTCCCA		
18_74756101_74757557_74845064_74846657_RR	0.003489147	0.089578901	AGATGTGTAAGTCACCAGGGAGTGCAATTCGACCTCGCGGGAGGGGTGCCAGGCCGCAT		
PRKCQ_Site10_Site4_FR	0.004639159	0.093849223	GTAATGGTGCCATCATAGCTCAAGCTCCTCGATCTATGAGGAAATGCCCCCAGCCTCCCA		
PRKCQ_Site10_Site4_RR	0.007812066	0.108064059	AATACAAAAGGATGGTAATTTTGCAATTCGATCTATGAGGAAATGCCCCCAGCCTCCCA	PCT/GB2016/051894	
PRKCZ_Site8_Site9_FR	0.000560117	0.079228864	CCCTCTCTCTAAAAGGTCTCAACATCACTCGATGGTGGCGGAGGTGGCCGGCAGGGTTGG		
MTHFD1_Site5_Site1_RF	0.000404338	0.079228864	ATAATCTCTCCGGCACATAATAAGTATTCGAATCGGGGGGGTTCGGGCGTGGGTTTCAG		
NFAT_Site6_Site1_FF	0.000514351	0.079228864	TCTAAAGGGATTCCACTATATGTAGATTCGAGGGGCGTGTGCGCGGTGGCGGGGCCCG		
PRKCQ_Site11_Site4_RR	0.006796573	0.102494645	AACTTATGATCTAATCTTGAATGTCTGTCGATCTATGAGGAAATGCCCCCAGCCTCCCA		

Probes	NR_R_P.Value	NR_R_adj.P.Val	Probe sequence	
			60 mer	WO 2016/207647
TNFRSF1A_Site5_Site6_FF	0.011987094	0.126537326	GAGGTGGGCAGATCACGGGTCAGGGTATCGAGGCCCATCATCTGGCGGGGAGACGGGAGC	
18_74845064_74846657_74864266_74864995_RF	0.008686097	0.111746517	ACTGAATATGAAAAAATGTAAAAATTATCGACCTCGGGGAGGGGTGCCAGGCCGCAT	
PRKCQ_Site7_Site4_RR	0.011239245	0.123381356	GATTTATAGCAAAATTACAAAAATGAGTCGATCTATGAGGAAATGCCCCAGCCTCCCA	
PRKCZ_Site5_Site9_RR	0.002885944	0.086622849	ACCAAGAGTTGGACCCCTTTTGTATGTTGATGTTGCGGAGGTGGCCGGCAGGGTTGG	
MAL_Site4_Site2_FR	0.000818457	0.0792228864	TATATTGCTATCTACTAGCAAAGGATAAATCGAAGAGGTTTCAGGGCGGTGCCCCGCGCT	
PRKCQ_Site9_Site4_RR	0.003669785	0.090200643	ATCAGTAAGCTGGTCAGCTACCCATGAATCGATCTATGAGGAAATGCCCCAGCCTCCCA	
TNFRSF14_Site3_Site8_FR	0.000995361	0.0792228864	TGAAAAACAGTTCACTCTGAGTTTCAGTCTCGAAGATTAGGTAAAGTTGGGGACGCGGAGA	
IFNAR1_Site2_Site4_RR	0.004801376	0.093849223	GTGCAGAGCGAGAGCGGGGCAGAGCGGTGCAAACTGGGAGAATTTCATCTGAAATGATTA	
IL-21R_Site5_Site2_RR	0.034533931	0.199109911	GAGGCAGGCAGATCATGAGGTCAGGAGTTCGAGCCCTGGACCCCGAGCCTAATGAGG	
19_10326358_10327821_10368389_10370560_RR	0.000174676	0.0792228864	GCTCACTGCAACCTCCACCTCCAGGTTTCGGAACCTCCTGATAACTTCAGCATTAACAG	
19_55449062_55451429_55484960_55486708_RF	7.78E-05	0.0792228864	AGGGTCTTGCTATGTTGCCAGGCTGGCCCTCGAGATCAGCCTGGGGCAACACCGGTGAAAAC	
TLR1_Site4_Site7_FR	0.000969535	0.0792228864	TGTAATATAAGCATAGCTCACTGCAGCCCTCGAAGCATTTGTACGACATTCATCTTCTT	
IRF5_Site8_Site2_FF	0.000148986	0.0792228864	ACAGAGGAGCGAGGCCCGATCCTTACTTTCGAACTCCTGACCTCGTATCTGCCCCACCTC	
SPRED2_Site4_Site8_RF	0.018236449	0.149371667	GGGTTTCACCAITGTTAGCCAGGATGGTCTCGATCTCCTGACCTCATGATCCGCTGCCTC	
IKKB_Site5_Site8_FR	0.013123191	0.130076121	GCATTTACCATGTTGGTGAGGCTGGTCTCGAAGAGTTTCACACGTGTCCAAATTTGGTGG	
TLR1_Site9_Site2_FF	0.002914123	0.086622849	CTGGGATCACAGGCATGTGCCACCATGCTCGACAAGAATAGTCTCCTTGTCTTGAACAT	
CD28_Site1_Site9_RR	0.003257956	0.088621062	GTATTTCTGGTTCTAGATCCTTGAGGAATCGAGCAGAGGAGTCTCTCCTGAGGCCACC	
12_10289678_10290500_10350455_10351677_RF	0.001491578	0.079419805	CGAGGGGGGGGATCACGAGGTCAGGAGATCGACCCCGACGTTCTCACCACCTGTTTCTT	
CD28_Site1_Site8_RR	0.007644106	0.107723492	GTATTTCTGGTTCTAGATCCTTGAGGAATCGACCTCCTCGGCTCAACCTATCCTCCACC	PCT/GB2016/051894

WO 2016/207647				PCT/GB2016/051894			
Probes	NR_R_P.Value	NR_R_adj.P.Val	60 mer	Probe sequence			
CXCL8_Site2_Site6_RF	0.002891692	0.086622849	GGGTTTCACTGTTAGCCAGGATGGTCTCGACCTCCCTGGCTCAAGTGATCTTCCCACC				
IL-23R_Site4_Site3_RF	0.001588257	0.079419805	TGCCCTAGAGATCTGTGGAACTTTGAACTCGATATATGAAAATAGTTTTTAATTATAAA				
RBPJ_Site14_Site13_FF	0.010539749	0.118804917	GGTGGGGGAATCACTTGAGGTCAGAAGTTCGAGACCATCCTGGGCAACATGGTAAAAACCC				
CHUK_Site7_Site2_RF	0.000132328	0.079228864	AATGGCACGATCACGGCTCACTGCAGCCTCGAATGTTACTGACAGTGGACACAGTAAGAA				
SH2B3_Site6_Site5_FF	0.003743845	0.090200643	GAGTTTGGCCATGTTGCCAGGCTGGTCTCGAGAACAGCCTGGCCAAACATGGTGAAACCC				
IRAK3_Site7_Site5_FR	0.00056928	0.079228864	AGGTCTCACTATGTTGCCCGGGCTGGTCTCGACGCCGAGGAGCTCTGCAGTGGGGGGCGTA				
CD28_Site4_Site2_RF	0.014801185	0.136839161	GGGTTTCAACCATGTTGGCGAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCGCCTGCC				
CD28_Site5_Site6_FR	0.007402719	0.106291976	GGTGGGTGGATCACCTGAGGTGAGGAGTTCGACCTAAGGGTGGTCATAATTCTGCTGCTG				
19_39424583_39425930_39445791_39449626_FF	0.001743055	0.079577656	GGGTCTCACAGCCTTCAGAGCTGAGAGCCTAGGCTTCAGTGAGCCATAATCACGCCCACTA				
IL-1a_and_IL-1b_Site1_Site7_RF	0.002815998	0.086622849	CTTTGGGAGGCCAAGGTGAGTGGATTGCTCGACATCTCATTTGATAGGATTAAGTCAACG				
IRAK3_Site7_Site1_FF	0.00166033	0.079419805	AGGTCTCACTATGTTGCCCGGGCTGGTCTCGAACAGCAGCGTGTGCGCCGACAGCGCGCC				
C5orf30_Site2_Site8_FR	0.00524841	0.094393734	TCTGTGCCCCAGGTTGGAGTACAGTGGCTCGAGGATGTCTCTATTTTGCACCTTATCTAA				
CXCL13_Site1_Site3_RR	6.56394E-05	0.079228864	TTATATCTCTACCTCCAAGCCTGGCAGTCGATTCCAAAGTGAAGCAAAAAAAACTTC				
14_55507409_55508411_55583475_55586339_RF	0.003368236	0.088703855	AAAGACCCTGTCTCTAAATAAATAGAACATCGAGATCATGCCACTGCACCTCCAGCCTGGG				
14_91450408_91451505_91524833_91527062_FF	0.004287708	0.093190996	GGGGTTTTTCCATGTTAGTCAGGCTGGTCTAATGGCTCCCTTACCTTGGCTGGCTGGGC				
IL-23_Site4_Site5_FR	0.021765214	0.160960834	AGTGGCATGATCACAGCTCACTGCCACCTCGAAACCAACCCCTGTGACTTCAACACCCAA				
IL-17A_Site3_Site1_RR	0.009698852	0.115042065	CCCTCCCTCAACATGCAGGGATTACAATTCTGAAGATGGTCTGAAGGAAGCAATTGGGAAA				

Example 1 - Table 8b. Stratifying between RA-MTX responders and non-responders

Probe Location				4 kb Sequence Location					
Chr	Start1	End1	Start2	End2	Chr	Start1	End1	Start2	End2
1	2460436	2460465	2486982		1	2487011	2456466	2460465	2486982
1	2457910	2457939	2486982		1	2487011	2457910	2461909	2486982
12	6443253	6443282	6472689		12	6472718	6439283	6443282	6472689
12	6452140	6452169	6472689		12	6472718	6448170	6452169	6472689
1	2460436	2460465	2539015		1	2539044	2456466	2460465	2539015
21	36117642	36117671	36260589		21	36260618	36117642	36121641	36260589
1	2486982	2487011	2540813		1	2540842	2486982	2490981	2536843
18	74845065	74845094	74866978		18	74867007	74845065	74849064	74867007
1	1977899	1977928	2066129		1	2066158	1973929	1977928	2066129
21	36206580	36206609	36260589		21	36260618	36206580	36210579	36260589
10	6520005	6520034	6577853		10	6577882	6516035	6520034	6577853
18	74756102	74756131	74845065		18	74845094	74756102	74760101	74849064
10	6454073	6454102	6577853		10	6577882	6450103	6454102	6577853
10	6448929	6448958	6577853		10	6577882	6448929	6452928	6577853
1	1977899	1977928	2125692		1	2125721	1973929	1977928	2125692
14	64856944	64856973	64805460		14	64805493	64852973	64856973	64805460
18	77135881	77135910	77156058		18	77156087	77131911	77135910	77152088
									77156087

Example 1 - Table 8b. Stratifying between RA-MTX responders and non-responders

Chr	Start1	End1	Start2	End2	Chr	Start1	End1	Start2	End2
10	6391740	6391769	6577853	6577882	10	6391740	6395739	6577853	6581852
12	6473688	6473717	6494374	6494403	12	6469718	6473717	6490404	6494403
18	74845065	74845094	74864966	74864995	18	74845065	74849064	74860996	74864995
10	6515356	6515385	6577853	6577882	10	6515356	6519355	6577853	6581852
1	2035712	2035741	2125692	2125721	1	2035712	2039711	2125692	2129691
2	95655674	95655703	95691307	95691336	2	95651704	95655703	95691307	95695306
10	6427823	6427852	6577853	6577882	10	6427823	6431822	6577853	6581852
1	2483531	2483560	2539015	2539044	1	2479561	2483560	2539015	2543014
21	34696685	34696714	34746263	34746292	21	34696685	34700684	34746263	34750262
16	27367634	27367663	27460580	27460609	16	27367634	27371633	27460580	27464579
19	10326359	10326388	10368390	10368419	19	10326359	10330358	10368390	10372389
19	55449063	55449092	55486679	55486708	19	55449063	55453062	55482709	55486708
4	38794092	38794121	38904213	38904242	4	38790122	38794121	38904213	38908212
7	128578517	128578546	128592079	128592108	7	128574547	128578546	128588109	128592108
2	65604070	65604099	65634253	65634282	2	65604070	65608069	65630283	65634282
8	42092338	42092367	42202562	42202591	8	42088368	42092367	42202562	42206561
4	38788263	38788292	38859677	38859706	4	38784293	38788292	38855707	38859706
2	204566973	204567002	204624489	204624518	2	204566973	204570972	204624489	204628488

Example 1 - Table 8b. Stratifying between RA-MTX responders and non-responders

Chr	Start1	End1	Start2	End2	Chr	Start1	End1	Start2	End2
12	10289679	10289708	10351648	10351677	12	10289679	10293678	10347678	10351677
2	204566973	204567002	204645538	204645567	2	204566973	204570972	204645538	204649537
4	74601393	74601422	74662726	74662755	4	74601393	74605392	74658756	74662755
1	67639374	67639403	67673763	67673792	1	67639374	67643373	67669793	67673792
4	26109288	26109317	26147759	26147788	4	26105318	26109317	26143789	26147788
10	101933094	101933123	101989686	101989715	10	101933094	101937093	101985716	101989715
12	111834072	111834101	111901271	111901300	12	111830102	111834101	111897301	111901300
12	66544383	66544412	66696510	66696539	12	66540413	66544412	66696510	66700509
2	204522870	204522899	204607547	204607576	2	204522870	204526869	204603577	204607576
2	204541606	204541635	204582161	204582190	2	204537636	204541635	204582161	204586160
19	39425901	39425930	39449597	39449626	19	39421931	39425930	39445627	39449626
2	113627760	113627789	113530289	113530318	2	113623789	113627789	113530289	113526289
12	66544383	66544412	66583104	66583133	12	66540413	66544412	66579134	66583133
5	102618306	102618335	102629447	102629476	5	102614336	102618335	102629447	102633446
4	78431568	78431597	78523781	78523810	4	78431568	78435567	78523781	78527780
14	55507410	55507439	55586310	55586339	14	55507410	55511409	55582340	55586339
14	91451476	91451505	91527033	91527062	14	91447506	91451505	91523063	91527062
12	56741028	56741057	56754855	56754884	12	56737058	56741057	56754855	56758854
6	52026497	52026526	52049432	52049461	6	52026497	52030496	52049432	52053431

Example 1A - Table 9a. RA vs. healthy (HC)

probe	GeneLocus	Probe_Count_Total	Probe_Count_Sig	HyperG_Stats	FDR_HyperG	Percent_Sig	reps..
3_112025276_112034935_112084448_112086795_RR	CD200	26	8	0.0009	0.034641	30.77	4
7_80168823_80173631_80193869_80200362_FF	CD36	127	31	3.33E-08	1.03E-05	24.41	2
10_98399260_98400639_98464393_98468588_RR	PIK3AP1	210	32	0.000597	0.026269	15.24	4
10_98397707_98399014_98464393_98468588_RR	PIK3AP1	210	32	0.000597	0.026269	15.24	2
10_98426247_98429729_98464393_98468588_RR	PIK3AP1	210	32	0.000597	0.026269	15.24	4
5_73482279_7353422_7459585_7461017_RR	ADCY2	364	41	0.028605	0.534587	11.26	4
1_167474157_167477896_167516923_167519477_FF	CD247	254	28	0.076338	0.800928	11.02	4
10_98413942_98416630_98464393_98468588_RR	PIK3AP1	210	32	0.000597	0.026269	15.24	3
10_98406449_98407502_98464393_98468588_RR	PIK3AP1	210	32	0.000597	0.026269	15.24	2
10_98374146_98380277_98464393_98468588_FR	PIK3AP1	210	32	0.000597	0.026269	15.24	4
10_98397707_98399014_98464393_98468588_FR	PIK3AP1	210	32	0.000597	0.026269	15.24	3
22_40991346_40993921_41008883_41010718_FR	MKL1	183	29	0.000555	0.026269	15.85	3
5_7375991_7381724_7459585_7461017_RF	ADCY2	364	41	0.028605	0.534587	11.26	4
22_40896154_40899434_41056322_41063897_FF	MKL1	183	29	0.000555	0.026269	15.85	2
10_98442806_98446178_98464393_98468588_FR	PIK3AP1	210	32	0.000597	0.026269	15.24	3
10_98464393_98468588_98520690_98524157_RF	PIK3AP1	210	32	0.000597	0.026269	15.24	4
10_98362077_98370186_98464393_98468588_FR	PIK3AP1	210	32	0.000597	0.026269	15.24	4
3_112025276_112034935_112094416_112098885_RR	CD200	26	8	0.0009	0.034641	30.77	4
5_7402050_7407728_7612925_7619203_RR	ADCY2	364	41	0.028605	0.534587	11.26	4
10_98362077_98370186_98464393_98468588_RR	PIK3AP1	210	32	0.000597	0.026269	15.24	3
11_93832833_93843526_93895630_93897747_FR	PANX1	29	5	0.088438	0.801142	17.24	2
10_98413942_98416630_98464393_98468588_FR	PIK3AP1	210	32	0.000597	0.026269	15.24	4
11_93843526_93849067_93895630_93897747_RR	PANX1	29	5	0.088438	0.801142	17.24	4
10_98442806_98446178_98464393_98468588_RR	PIK3AP1	210	32	0.000597	0.026269	15.24	4
22_40871339_40876622_41008883_41010718_FR	MKL1	183	29	0.000555	0.026269	15.85	4
22_40848625_40853672_41008883_41010718_FR	MKL1	183	29	0.000555	0.026269	15.85	4

Example 1A - Table 9a. RA vs. healthy (HC)

Probe	GeneLocus	Probe_Count_Total	Probe_Count_Sig	HyperG_Stats	FDR_HyperG	Percent_Sig	reps..
17_73322245_73323380_73394039_73395972_FF	GRB2	270	32	0.026522	0.534587	11.85	4
2_173587206_173590304_173788215_173791519_FF	RAPGEF4	195	22	0.088283	0.801142	11.28	3
X_19753406_19760963_19778202_19779729_RF	SH3KBP1	291	33	0.042262	0.681595	11.34	4
22_40796444_40801147_40909402_40912220_RR	MKL1	183	29	0.000555	0.026269	15.85	4
22_40896154_40899434_40931576_40935727_FF	MKL1	183	29	0.000555	0.026269	15.85	4
22_40796444_40801147_40871339_40876622_RR	MKL1	183	29	0.000555	0.026269	15.85	4
7_55061795_55064635_55224588_55235839_RR	EGFR	209	35	5.02E-05	0.007733	16.75	4
11_119100257_119101910_119157901_119160975_FR	CBL	41	7	0.050372	0.71046	17.07	3
1_167399005_167402982_167413430_167415364_RF	CD247	254	28	0.076338	0.800928	11.02	4
22_40909402_40912220_40931576_40935727_RF	MKL1	183	29	0.000555	0.026269	15.85	4
17_73352521_73353799_73428595_73430537_RF	GRB2	270	32	0.026522	0.534587	11.85	4
7_80168823_80173631_80308967_80317006_FF	CD36	127	31	3.33E-08	1.03E-05	24.41	4
22_40909402_40912220_41075227_41079714_RF	MKL1	183	29	0.000555	0.026269	15.85	4
X_19545819_19548298_19747473_19749276_FF	SH3KBP1	291	33	0.042262	0.681595	11.34	4
10_98401433_98405814_98464393_98468588_RR	PIK3AP1	210	32	0.000597	0.026269	15.24	4
17_73347837_73349062_73428595_73430537_RR	GRB2	270	32	0.026522	0.534587	11.85	4
X_19644496_19650796_19920428_19925492_FF	SH3KBP1	291	33	0.042262	0.681595	11.34	2
7_80078955_80088693_80121443_80124810_RR	CD36	127	31	3.33E-08	1.03E-05	24.41	4
22_40871339_40876622_40909402_40912220_RR	MKL1	183	29	0.000555	0.026269	15.85	4
22_40888137_40890603_41008883_41010718_RR	MKL1	183	29	0.000555	0.026269	15.85	2
11_93858215_93861587_93895630_93897747_FR	PANX1	29	5	0.088438	0.801142	17.24	4
22_40909402_40912220_41008883_41010718_RF	MKL1	183	29	0.000555	0.026269	15.85	2
22_40909402_40912220_40944160_40947074_RF	MKL1	183	29	0.000555	0.026269	15.85	3
17_73401174_73403644_73443323_73445724_FF	GRB2	270	32	0.026522	0.534587	11.85	4

Example 1A - Table 9b. RA vs. healthy (HC)

Avg CV	logFC	AveExpr	t	P.Value	adj.P.Val	B	FC	FC 1	LS.x	Loop detected
22.5672	0.685577	0.685577	7.208757	6.20E-05	0.067791	2.137374	1.608346	1.608346	1	RA
2.8578	0.513592	0.513592	5.758269	0.000319	0.067791	0.70324	1.4276	1.4276	1	RA
3.9824	0.509022	0.509022	5.490428	0.000444	0.067791	0.402326	1.423085	1.423085	1	RA
4.8484	0.49927	0.49927	5.744216	0.000324	0.067791	0.687747	1.413498	1.413498	1	RA
3.7034	0.497429	0.497429	5.438056	0.000474	0.067791	0.342088	1.411696	1.411696	1	RA
25.3272	0.481183	0.481183	5.488293	0.000445	0.067791	0.39988	1.395888	1.395888	1	RA
3.902	0.477912	0.477912	4.775775	0.001128	0.084274	0.459904	1.392726	1.392726	1	RA
3.4802	0.473702	0.473702	5.498122	0.000439	0.067791	0.411136	1.388668	1.388668	1	RA
4.5316	0.466456	0.466456	5.514684	0.00043	0.067791	0.430069	1.381711	1.381711	1	RA
4.3162	0.464731	0.464731	4.947314	0.000895	0.079113	0.244976	1.38006	1.38006	1	RA
4.07	0.450917	0.450917	5.861317	0.000281	0.067791	0.815863	1.366909	1.366909	1	RA
2.468	0.449615	0.449615	6.187618	0.000191	0.067791	1.161199	1.365676	1.365676	1	RA
28.4296	0.44389	0.44389	6.177656	0.000193	0.067791	1.150906	1.360267	1.360267	1	RA
1.959	0.434213	0.434213	5.114367	0.000718	0.076048	-0.04052	1.351174	1.351174	1	RA
4.0442	0.433359	0.433359	5.322391	0.000549	0.070404	0.207413	1.350374	1.350374	1	RA
4.3738	0.433217	0.433217	4.906707	0.000945	0.079113	0.295398	1.350241	1.350241	1	RA
3.6068	0.431233	0.431233	5.069327	0.000762	0.078192	0.095173	1.348386	1.348386	1	RA
7.675	0.430701	0.430701	3.417531	0.008129	0.14654	2.330752	1.347889	1.347889	1	RA
2.4692	0.429497	0.429497	4.529234	0.001583	0.094398	0.777617	1.346764	1.346764	1	RA
6.9092	0.424613	0.424613	4.919181	0.00093	0.079113	0.279879	1.342212	1.342212	1	RA
0.861	0.418444	0.418444	3.453029	0.007695	0.144805	2.278502	1.336486	1.336486	1	RA
3.8908	0.418119	0.418119	4.618943	0.001398	0.092869	0.660814	1.336184	1.336184	1	RA
3.9704	0.412721	0.412721	3.265842	0.01029	0.157013	-2.55552	1.331194	1.331194	1	RA
4.374	0.411793	0.411793	4.953972	0.000888	0.079113	0.236736	1.330338	1.330338	1	RA
2.2892	0.399862	0.399862	5.585453	0.000394	0.067791	0.510449	1.319382	1.319382	1	RA

Example 1A - Table 9b. RA vs. healthy (HC)

Avg CV	logFC	AveExpr	t	P.Value	adj.P.Val	B	FC	FC 1	LS.x	Loop detected
3.2674	0.398739	0.398739	5.342329	0.000535	0.06998	0.230789	1.318355	1.318355	1	RA
4.4132	0.397225	0.397225	4.054922	0.003115	0.118804	-1.417412	1.316972	1.316972	1	RA
4.4052	0.395013	0.395013	5.008	0.000826	0.079113	-0.170148	1.314955	1.314955	1	RA
29.626	0.393804	0.393804	3.646302	0.005726	0.137876	-1.996639	1.313853	1.313853	1	RA
2.9968	0.388334	0.388334	3.462029	0.007589	0.144805	-2.265278	1.308881	1.308881	1	RA
3.9034	0.386289	0.386289	4.286116	0.00223	0.107797	-1.100959	1.307027	1.307027	1	RA
4.137	0.385157	0.385157	3.466476	0.007538	0.144805	-2.258746	1.306002	1.306002	1	RA
16.4644	0.3851	0.3851	2.766312	0.022719	0.20414	-3.307709	1.30595	1.30595	1	RA
4.0986	0.385001	0.385001	5.604724	0.000385	0.067791	0.532193	1.305861	1.305861	1	RA
16.621	0.384436	0.384436	2.93912	0.017235	0.186238	-3.046125	1.305349	1.305349	1	RA
2.9874	0.37983	0.37983	3.321482	0.009435	0.152196	-2.472807	1.301188	1.301188	1	RA
3.5382	0.379663	0.379663	5.968229	0.000247	0.067791	0.930887	1.301038	1.301038	1	RA
26.1988	0.37908	0.37908	4.66095	0.001319	0.089846	-0.606589	1.300513	1.300513	1	RA
3.7712	0.37846	0.37846	3.732483	0.005026	0.129648	-1.872495	1.299954	1.299954	1	RA
5.4524	0.377994	0.377994	4.038461	0.003191	0.119364	-1.440267	1.299533	1.299533	1	RA
5.6888	0.377948	0.377948	4.703842	0.001244	0.08826	-0.551533	1.299492	1.299492	1	RA
5.8806	0.377857	0.377857	5.484525	0.000447	0.067791	0.39556	1.29941	1.29941	1	RA
1.8516	0.374725	0.374725	5.431708	0.000478	0.067791	0.334755	1.296592	1.296592	1	RA
3.655	0.367548	0.367548	2.983814	0.016052	0.184597	-2.978625	1.290159	1.290159	1	RA
2.5222	0.365318	0.365318	3.184975	0.01168	0.160965	-2.676233	1.288166	1.288166	1	RA
2.1288	0.364614	0.364614	5.288952	0.000573	0.070842	0.168055	1.287537	1.287537	1	RA
3.2306	0.363046	0.363046	3.338237	0.009192	0.151385	-2.447957	1.286139	1.286139	1	RA
5.2484	0.36296	0.36296	3.68495	0.0054	0.134402	-1.940843	1.286062	1.286062	1	RA
2.7764	0.362854	0.362854	3.558769	0.006542	0.144805	-2.123725	1.285967	1.285967	1	RA
2.8922	0.362136	0.362136	3.154388	0.012256	0.165143	-2.722033	1.285327	1.285327	1	RA

Example 1A - Table 9C. RA vs. healthy (HC)

Probe sequence/ Probe location		Chr	Start1	End1	Start2	End2	Chr	Start1	End1	Start2	End2
60 mer											
TATATAATTCACCTTTGTTTTTAATAATCGAA		3	112025277	112025306	112084449	112084478	3	112025277	1120292	1120844	112088
ACATAACTGTTCTAAAATATGTCAGT									76	49	448
TGCTGAAAGAAAAACACAATTATTTAAAGTCGA		7	80173602	80173631	80200333	80200362	7	80169632	8017363	8019636	802003
GACCATCTAGCTAACACGGTGAAACCC									1	3	62
GTTTTAACATTTAAAGATAAAATCCCATCGAA		10	98399261	98399290	98464394	98464423	10	98399261	9840326	9846439	984683
CCGAGGAGGCAGAGGTAGCAGTGAGC									0	4	93
AGCTGATTGTGTAACCTCTCAGTCTGAGCTCGAACCC		10	98397708	98397737	98464394	98464423	10	98397708	9840170	9846439	984683
AGGGAGGCAGAGGTAGCAGTGAGC									7	4	93
GGGAAATAAAATATTAAGAGCTTTAGTGTGGAACCC		10	98426248	98426277	98464394	98464423	10	98426248	9843024	9846439	984683
AGGGAGGCAGAGGTAGCAGTGAGC									7	4	93
TACCAGGAAGATATTTTATAAATGAATGTCGAAGACA		5	7348280	7348309	7459586	7459615	5	7348280	7352279	7459586	746358
GTTTGAGATTGCTTTCTAG									1674778	1675154	167519
TAAGTGGGAGAAAAGACAAAGATTCTCTCGAGGTG		1	167477867	167477896	167519448	167519477	1	167473897	96	78	477
AGCGGATCACCTGAGGTGAGGAGT									9841794	9846439	984683
AGATCTTAAAGCAAGCTAAAGAGCTATTCGAACCC		10	98413943	98413972	98464394	98464423	10	98413943	2	4	93
AGGGAGGCAGAGGTAGCAGTGAGC									9841044	9846439	984683
TCTCCTTTTGGGCACATAGGACATAAAAATCGAACC		10	98406450	98406479	98464394	98464423	10	98406450	9	4	93
CAGGGAGGCAGAGGTAGCAGTGAGC									9838027	9846439	984683
TTCATTCCCGCAAAAGGGTCATATATACTCGAACCC		10	98380248	98380277	98464394	98464423	10	98376278	7	4	93
AGGGAGGCAGAGGTAGCAGTGAGC									9839901	9846439	984683
ATACTGACACACTATCCACCACAAAGTCGAACCC		10	98398985	98399014	98464394	98464423	10	98395015	4	4	93
AGGGAGGCAGAGGTAGCAGTGAGC									4099392	4100888	410128
CTAATGTGCTAGTTGTCCACATATTAATCGAGC		22	40993892	40993921	41008884	41008913	22	40989922	1	4	83
CTGCAGTGAGCCATGATCATGCCACT											
TTCCTTTTAAAGCTTGTCTCTATCATTCGAGATA		5	7375992	7376021	7460988	7461017	5	7375992	7379991	7457018	746101
ATTTAGAATTAAAGGAATAAAC									4089943	4105989	410638
AGGTTTTCAGGTTGGCTGGGATGGTCTCGAGACC		22	40899405	40899434	41063868	41063897	22	40895435	4	8	97
AGCCTGACCAACATGGAGAAACCC									9844617	9846439	984683
GGAAACCAAACTGGAATTCAGGAGACAAATTCGAACCC		10	98446149	98446178	98464394	98464423	10	98442179	8	4	93
AGGGAGGCAGAGGTAGCAGTGAGC											

Example 1A - Table 9C. RA vs. healthy (HC)

60 mer	Chr	Start1	End1	Start2	End2	Chr	Start1	End1	Start2	End2
CACATTAACACCTGTCAATAACAGGATTCGAACCCA GGGAGGCAGAGGTAGCAGTGAGC	10	98464394	98464423	98524128	98524157	10	98464394	9846839	9852015	985241
GTACAAAGAAGTGATGTAGCATGTCTCTGTC GAACCCAGGAGGCAGAGGTAGCAGTGAGC	10	98370157	98370186	98464394	98464423	10	98366187	9837018	9846439	984683
TATATAATTCACATTTGTTTTTAATAATCG AAGGACATATGATGGGTGGCTCGCTCG	3	112025277	112025306	112094417	112094446	3	112025277	1120292	1120944	112098
AGAAATGAGTCAGGTTCAATGAATTGTCTC GAGACCATCATGGCTAACACGGTGAAACCC	5	7402051	7402080	7612926	7612955	5	7402051	7406050	7612926	761692
CAAGTGATGGGACACCCACCATGTCCTCG AACCAGGAGGCAGAGGTAGCAGTGAGC	10	98362078	98362107	98464394	98464423	10	98362078	9836607	9846439	984683
AATCTTCATGAGGAGGCAATCAAGATGTC GACTGCTGTGTAGCAATGAGCGAGGCTCC	11	93843497	93843526	93895631	93895660	11	93839527	9384352	9389563	938996
GAAATCACCGTCGGCAGGTTCTGCTGCTTC GAACCCAGGAGGCAGAGGTAGCAGTGAGC	10	98416601	98416630	98464394	98464423	10	98412631	9841663	9846439	984683
GTCAAACCTTTGAAAACTGCAGCTCCAGTCCG ACTGCTGTGCTAGCAATGAGCGAGGCTCC	11	93843527	93843556	93895631	93895660	11	93843527	9384752	9389563	938996
GTGTGACAATTTTCACAGAAGCGTTGTTCG AACCAGGAGGCAGAGGTAGCAGTGAGC	10	98442807	98442836	98464394	98464423	10	98442807	9844680	9846439	984683
AATGCTTATGTTCTAATTCCTCAAAAGGAATCG AGCTGCAGTGAGCCATGATGCCACT	22	40876593	40876622	41008884	41008913	22	40872623	4087662	4100888	410128
GCTCTGCAAGAAGACAGAGCAAGGTCTTCG AGCTGCAGTGAGCCATGATGCCACT	22	40853643	40853672	41008884	41008913	22	40849673	4085367	4100888	410128
GGGTTTACCGTGTAGCCAGGATGGTCTCG AGACCATCTGGCTAACATGGTGAAACCA	17	73323351	73323380	73395943	73395972	17	73319381	7332338	7339197	733959
ATATAAATTACATGTCAAGAAGATAATGTCTG AGACCATCTGACCAACATGGTGAAACCT	2	173590275	173590304	173791490	173791519	2	173586305	1735903	1737875	173791
CATGATAGTTAAGAGATCATATCTAGAATCG ATTCTCTATTTCAATTTATTTCCACTGTAA	X	19753407	19753436	19779700	19779729	X	19753407	1975740	1977573	197797
GGGTTTCAACATATTGGCCAGGCTGGTCTCG AGACCAGCCTGGCCAAACATGGTGAAACCC	22	40796445	40796474	40909403	40909432	22	40796445	4080044	4090940	409134
AGGTTTGGCAAGTTGGCTGGGATGGTCTCG AGACCATCTGGCCAAACATGGTGAAACCC	22	40899405	40899434	40935698	40935727	22	40895435	4089943	4093172	409357

Example 1A - Table 9C. RA vs. healthy (HC)

60 mer	Chr	Start1	End1	Start2	End2	Chr	Start1	End1	Start2	End2
GGCTGGCAGATCACCTAAGGTCAGGCAATTCG	11	119101881	119101910	119157902	119157931	11	119097911	1191019	1191579	119161
AGAGCATGAATAAAGACTGTGAAGGCT								10	02	901
GAGTGATTGGTTCCGAGGTCAGGAGGTC	7	55061796	55061825	55224589	55224618	7	55061796	5506579	5522458	552285
GACATATTCCTGTTCCCTTGGGAATAAAAA								5	9	88
TCAGGTACTTCTTAGCCTTATGGCTTCG	1	167399006	167399035	167415335	167415364	1	167399006	1674030	1674113	167415
ATGTGAGAGGCACCTCTTTCACTAATAG								05	65	364
GGTTTCACCATGTTGGCCAGGATGGTCTC	22	40909403	40909432	40935698	40935727	22	40909403	4091340	4093172	409357
GAGACGAGCTGGCCAAACATGGTGAACCC								2	8	27
TTTATATTTTAAAAATTTGGGTTTTTTTTTCG	17	73352522	73352551	73430508	73430537	17	73352522	7335652	7342653	734305
AGGCTGCAATGAGCCATGATCACACCACT								1	8	37
TGCTGAAAGAAAAACACAAATTTATTTAAGTC	7	80173602	80173631	80316977	80317006	7	80169632	8017363	8031300	803170
GAATAAATGTTGGCTATCTTACAGTGATT								1	7	06
GGGTTTCACCATGTTAGCCAGGATGGTCTC	22	40909403	40909432	41079685	41079714	22	40909403	4091340	4107571	410797
GAGACGAGCTGGCCAAACATGGTGAACCC								2	5	14
GAGTTTCACCATGTTGACCAAGGCTGGTCTC	X	19548269	19548298	19749247	19749276	X	19544299	1954829	1974527	197492
GAGATCAGCTGGGCAACATGGTGAACCC								8	7	76
GGGAGGACTGGATCAGGAATCTGTGCTTC	10	98401434	98401463	98464394	98464423	10	98401434	9840543	9846439	984683
GAACCCAGGAGGAGGAGGAGGAGCAGTGAGC								3	4	93
TACAACAAATTAAGATATCACCTATATTCTCGA	17	73347838	73347867	73428596	73428625	17	73347838	7335183	7342859	734325
GACCATCCTAGCTAACATGGTGAAATCT								7	6	95
GAGGGAAAAATACTAAGGCCACTAAAAATCG	X	19650767	19650796	19925463	19925492	X	19646797	1965079	1992149	199254
AGACCATCCTGGACAAACATGGAGAAACAC								6	3	92
GGGTTTAAACATATTGGCCAGGCTGGTCTCG	7	80078956	80078985	80121444	80121473	7	80078956	8008295	8012144	801254
AGACCAGCCTGGCCAAATGGTGAAACCC								5	4	43
GGATTTCACCATGTTGGCCAGGCTGGTCTCG	22	40871340	40871369	40909403	40909432	22	40871340	4087533	4090940	409134
AGACCAGCCTGGCCAAACATGGTGAAACCC								9	3	02
GTTTATTGCAGCATTTGGCTGTGGAGACTCG	22	40888138	40888167	41008884	41008913	22	40888138	4089213	4100888	410128
AGCTGCAGTGAGCCATGATCATGCCACT								7	4	83
AAACGGGACACAGCAGCGCTACTCAGGCCTC	11	93861558	93861587	93895631	93895660	11	93857588	9386158	9389563	938996
GACTGCTGTGTAGCAATGAGCGAGGCTCC								7	1	30
CCATGTTGGTCAGGCTGGTCTCAAACCTCTCGA	22	40909403	40909432	41010689	41010718	22	40909403	4091340	4100671	410107
GACCAGCCTGGCCAAACATGGTGAAACCC								2	9	18

Example 1A - Table 9C. RA vs. healthy (HC)

60 mer	Chr	Start1	End1	Start2	End2	Ch r	Start1	End1	Start2	End2
GGGTTTCTCCATGCTGGTCAGGCTGGTCTCG	22	40909403	40909432	40947045	40947074	22	40909403	4091340	4094307	409470
AGACCAGCCTGGCCAAACATGGTGAAACCC										74
GGGTTTCGCCATGTTGGCCAGGCTGGTCTCG	17	73403615	73403644	73445695	73445724	17	73399645	7340364	7344172	734457
AGACCAGCCTGGCCAAACATGGTGAAACCC										24

CLAIMS

1. A method of determining responsiveness to a specific therapy for rheumatoid arthritis in a subject, comprising detecting the presence or absence of 5 or more chromosomal interactions from Table 7a or Table 8a, wherein the specific therapy for rheumatoid arthritis comprises methotrexate or a pharmaceutically acceptable salt thereof.
2. A method according to claim 1, wherein 5 to 500 chromosomal interactions are typed.
3. A method according to claim 2 wherein 20 to 300 chromosomal interactions are typed.
4. A method according to claim 3 wherein 50 to 100 chromosomal interactions are typed.

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:

PCT/GB2016/051894

(22) International Filing Date:

24 June 2016 (24.06.2016)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

1511079.4	24 June 2015 (24.06.2015)	GB
1511080.2	24 June 2015 (24.06.2015)	GB
1519555.5	5 November 2015 (05.11.2015)	GB

(71) Applicant: **OXFORD BIODYNAMICS LIMITED**
[GB/GB]; 26 Beaumont Street, Oxford Oxfordshire OX1 2NP (GB).(72) Inventors: **RAMADASS, Aroul**; Oxford BioDynamics Limited, 26 Beaumont Street, Oxford Oxfordshire OX1 2NP (GB). **HUNTER, Ewan**; Oxford BioDynamics Limited, 26 Beaumont Street, Oxford Oxfordshire OX1 2NP (GB). **AKOULITCHEV, Alexandre**; Oxford BioDynamics Limited, 26 Beaumont Street, Oxford Oxfordshire OX1 2NP (GB).(74) Agent: **AVIDITY IP**; Broers Building, Hauser Forum, 21 J J Thomson Ave, Cambridge Cambridgeshire CB3 0FA (GB).

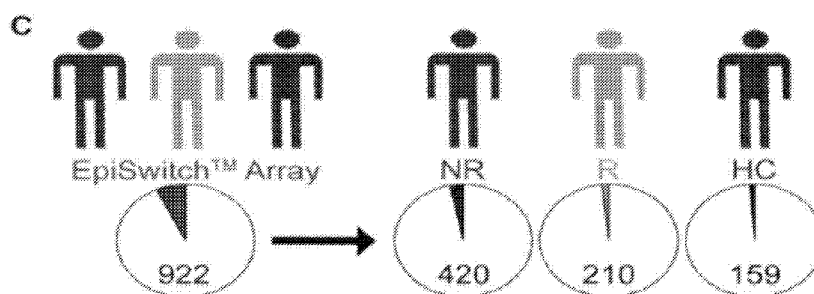
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: EPIGENETIC CHROMOSOME INTERACTIONS



(57) Abstract: A method of determining responsiveness to therapy for rheumatoid arthritis.

FIGURE 1

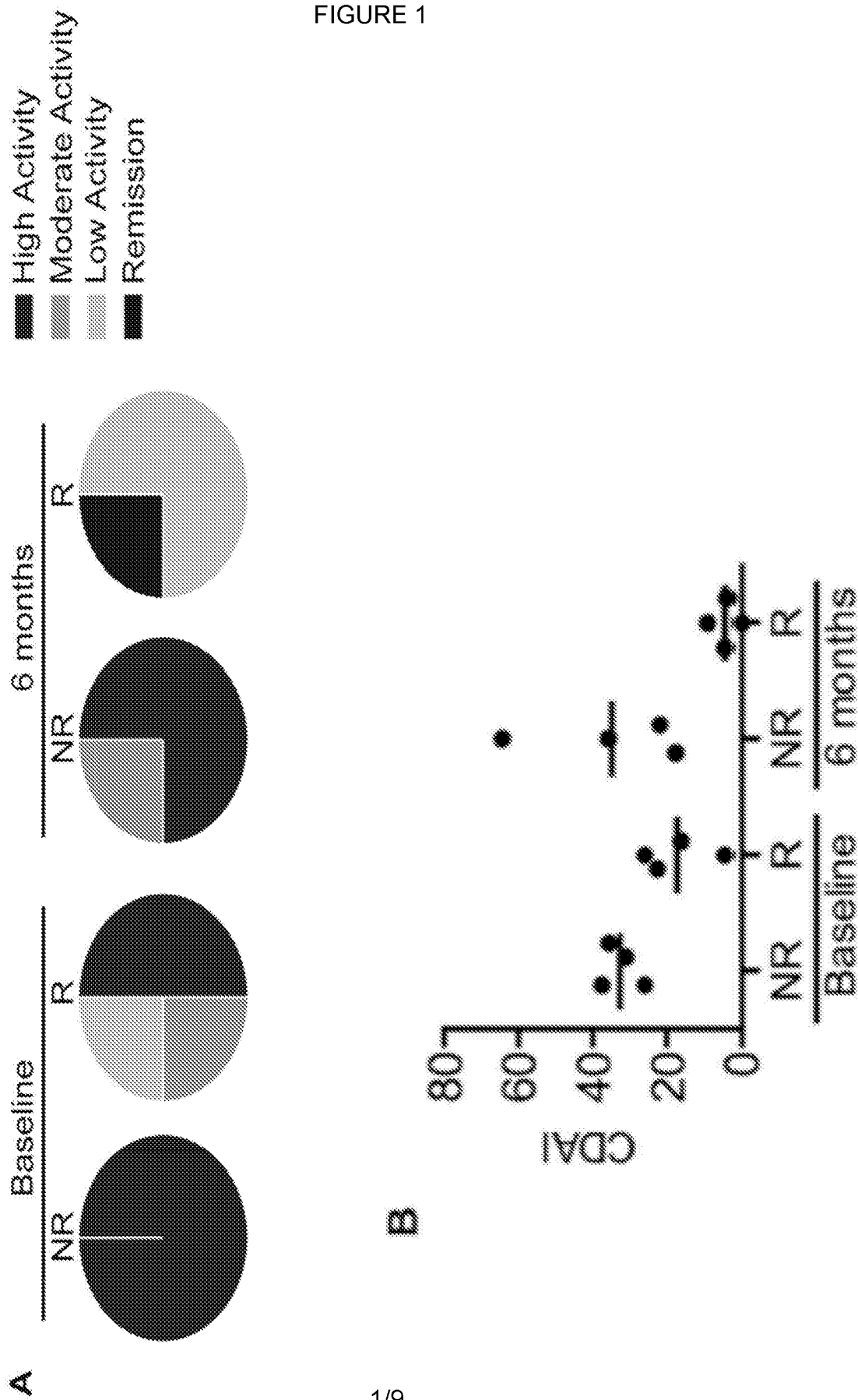


FIGURE 1

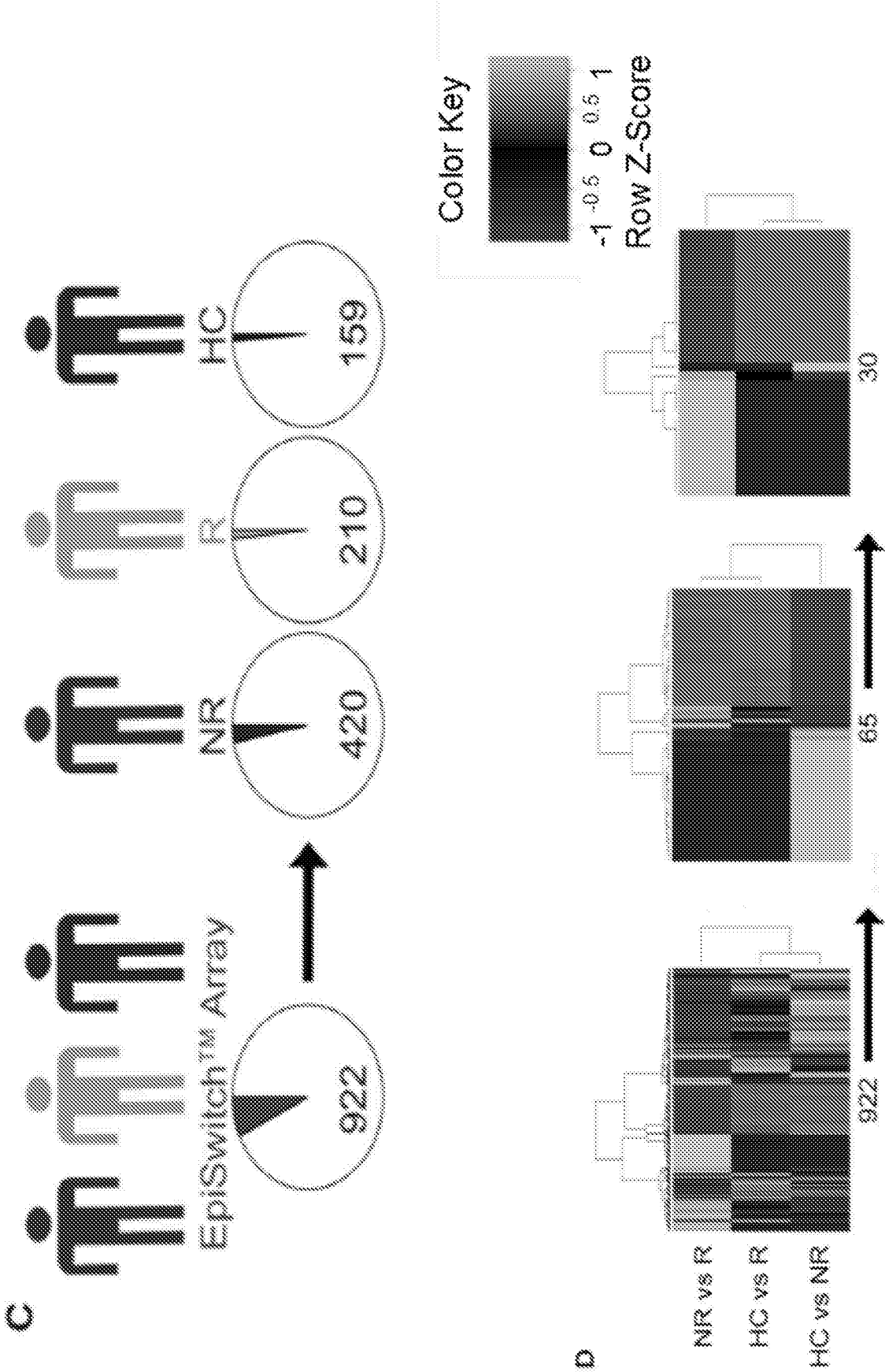


FIGURE 2

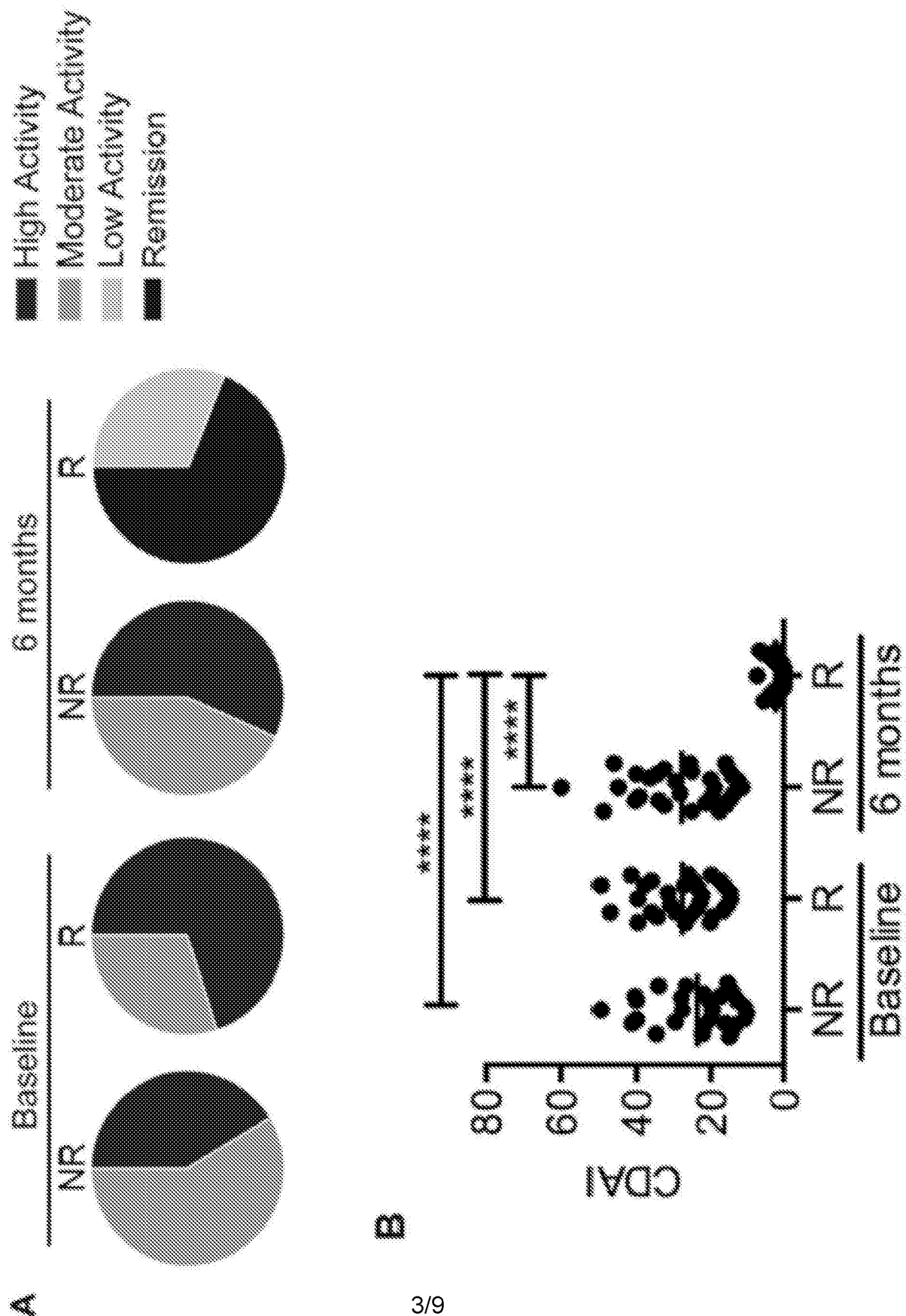


FIGURE 2

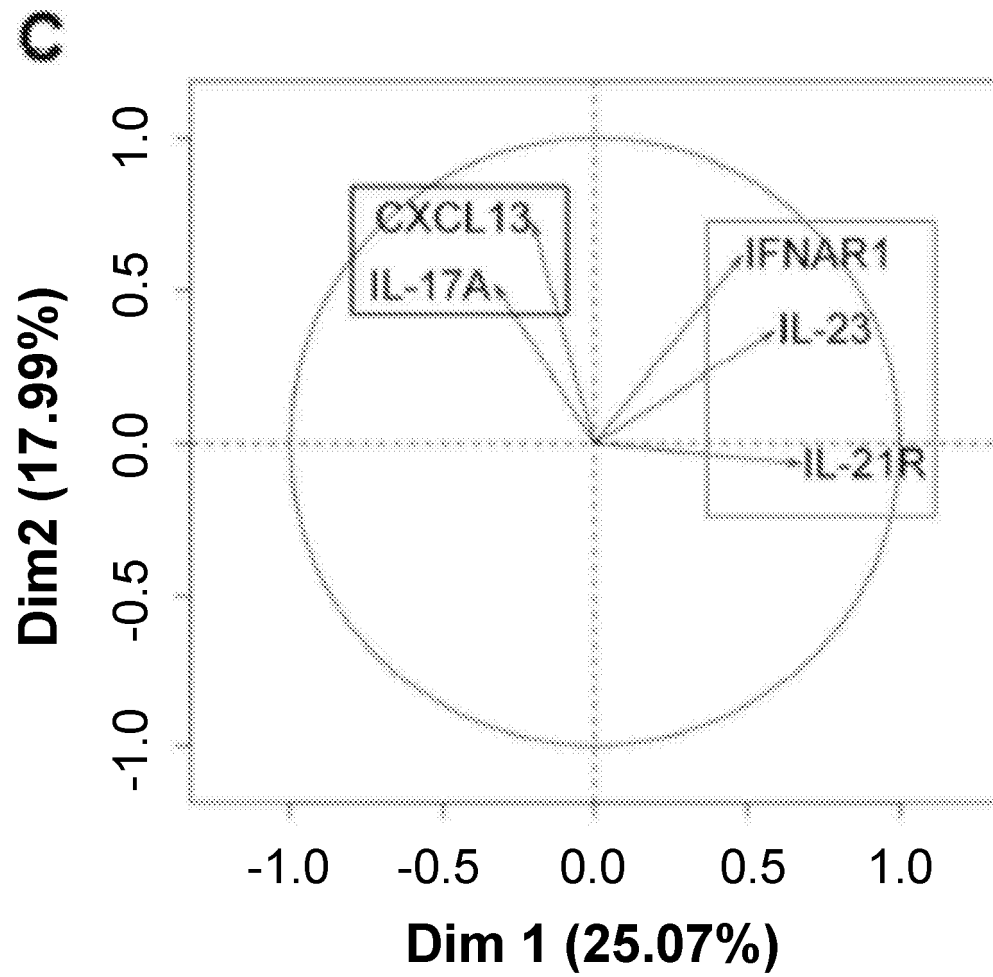


FIGURE 2

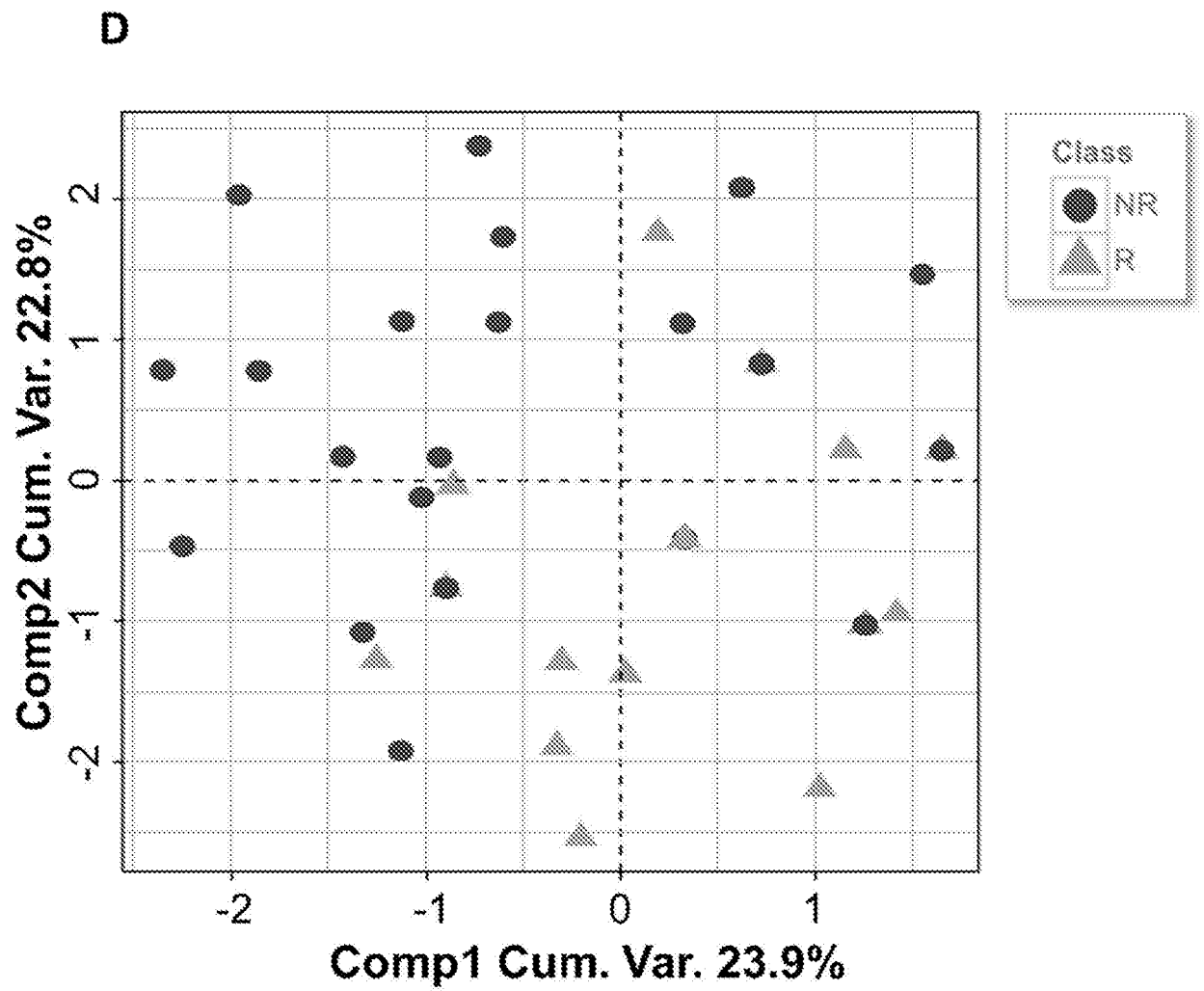


FIGURE 3

A

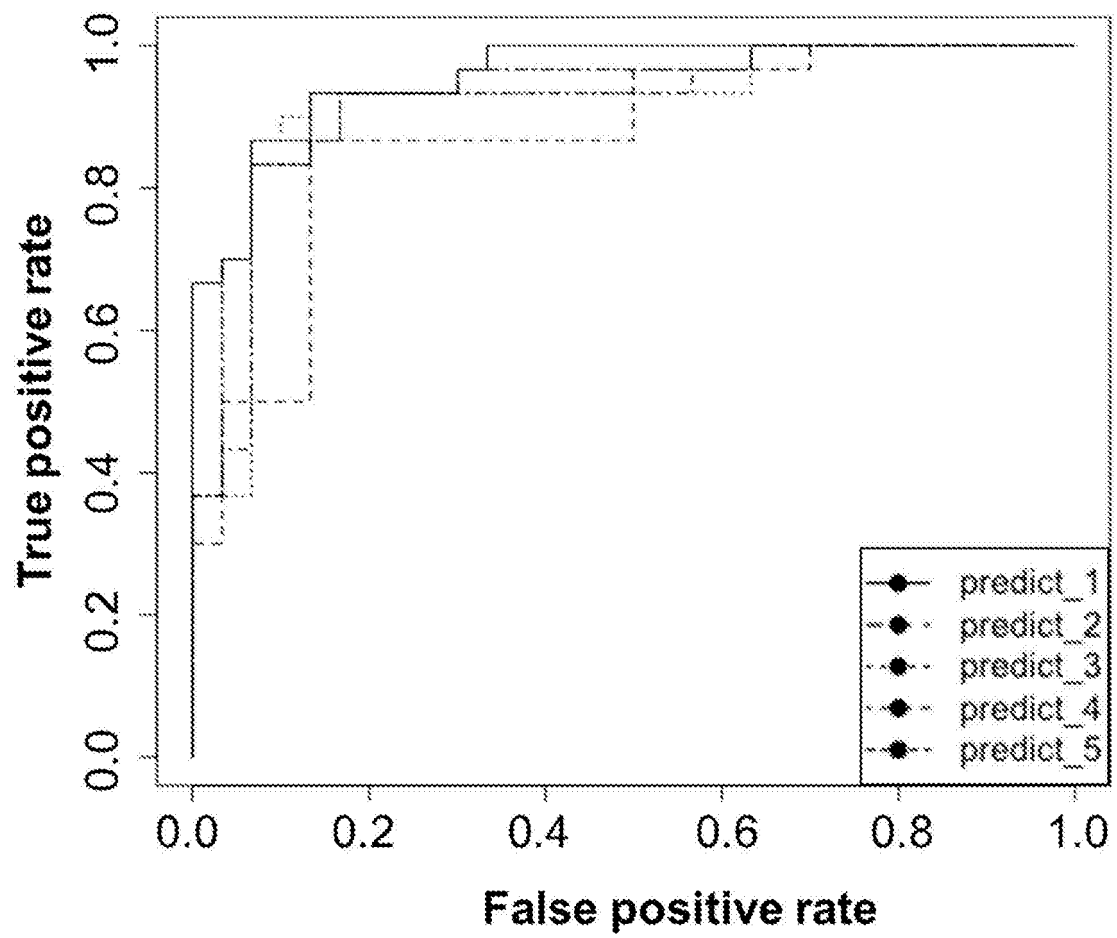


FIGURE 3

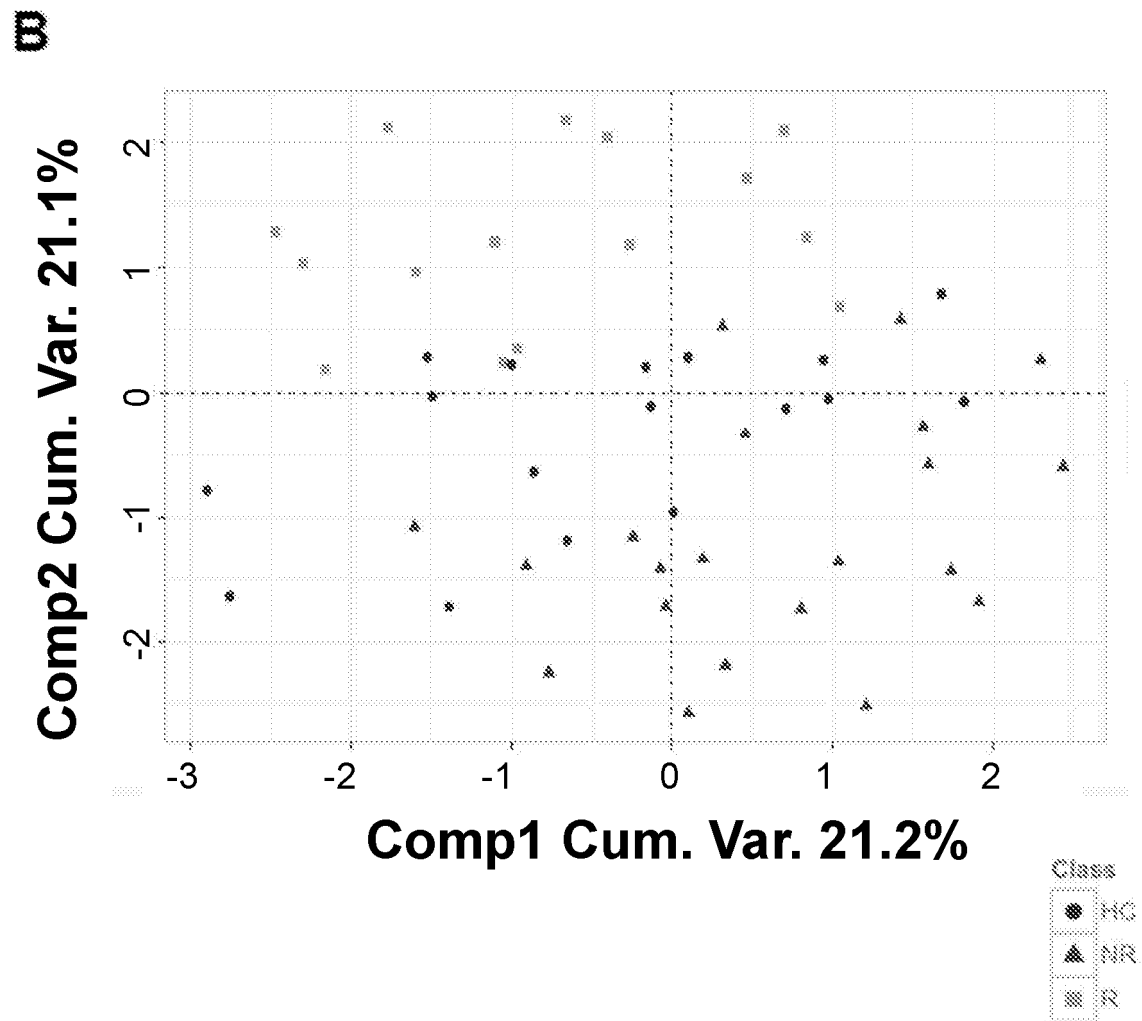


FIGURE 4

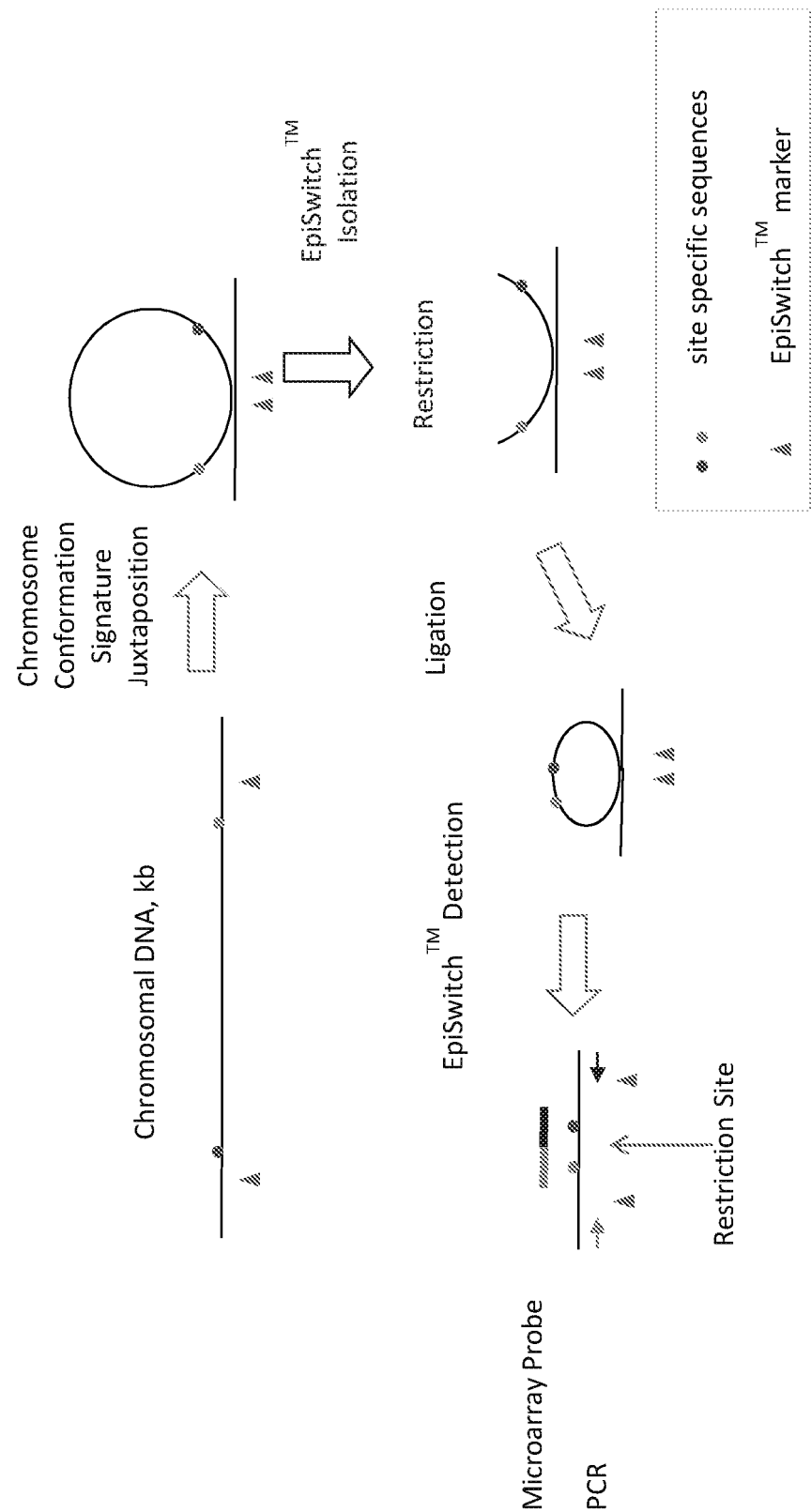


FIGURE 5

