

(51) International Patent Classification:
C12Q 1/68 (2006.01)(21) International Application Number:
PCT/GB2016/051097(22) International Filing Date:
21 April 2016 (21.04.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1506806.7 22 April 2015 (22.04.2015) GB

(71) Applicant: THE UNIVERSITY COURT OF THE UNIVERSITY OF EDINBURGH [GB/GB]; Old College, South Bridge, Edinburgh, Midlothian EH8 9YL (GB).

(72) Inventors: GRAY, Mohini; MRC Centre for Inflammation Research University of Edinburgh, The Queen's Medical Research Institute, 47 Little France Crescent, Edinburgh Midlothian EH16 4TJ (GB). COWAN, Graeme; Institute of Immunology and Infection Research University of Edinburgh, Ashworth Labs, Kings Buildings, West Mains Road, Edinburgh Midlothian EH9 3JT (GB).

(74) Agent: HGF LIMITED (GLASGOW); 1 City Walk, Leeds, Yorkshire LS11 9DX (GB).

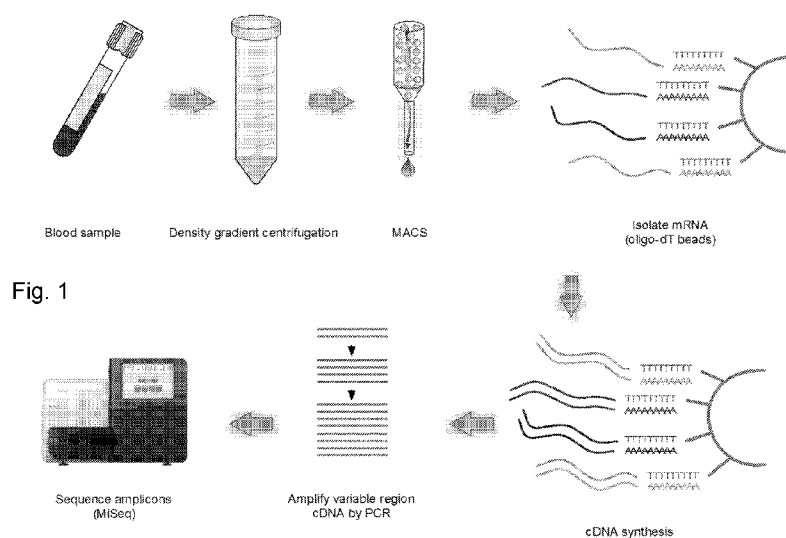
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, 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))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: RHEUMATOID ARTHRITIS PATIENT STRATIFICATION



(57) Abstract: The present invention relates to methods of identifying those patients with rheumatoid arthritis who are unlikely to respond to treatment with conventional disease modifying anti-rheumatic drugs (DMARDs). This allows for patient stratification and early selection of the most appropriate treatment for the subject. The method involves determining B cell hypermutation rates and using this as an indicator of patient responsiveness to conventional DMARDs. Associated assays and kits are also provided.



Rheumatoid Arthritis Patient Stratification

The present invention relates to methods of identifying those patients with rheumatoid arthritis who are unlikely to respond to treatment with conventional disease modifying anti-rheumatic drugs (DMARDs). This allows for patient stratification and early selection of the most appropriate treatment for the subject.

Background of the Invention

Rheumatoid Arthritis (RA) is the commonest, incurable, autoimmune arthritis, affecting 1% of the world's population, with over 400,000 new cases in Europe and the USA each year. A delay in controlling joint inflammation by even a few months with DMARDs leads to a poorer long-term outcome. Currently both the direct and indirect costs of managing RA in Europe alone amount to over £25Bn per year. Whilst there is no cure, early aggressive control of inflammation can reduce the long term risk of joint destruction, disability and premature death.

Methotrexate (MTX) (as monotherapy or in combination with other conventional DMARDS such as hydroxychloroquine, sulphasalazine and leflunomide) is the anchor DMARD in the management of RA due to its efficacy, safety and low cost (1). However a significant minority of patients are unresponsive to this 'gold-standard' therapy whom we refer to as primary-DMARD-Non-Responders [1°-DNRs] and are likely the same subset of RA patients previously referred to as multi-drug resistant (2). Patients must fail multiple DMARDs to be eligible for the more expensive (but efficacious) biologic therapies, such as anti-TNF α , anti IL-6 or rituximab (biologic DMARDs); meanwhile their disease progresses relentlessly, potentially leading to permanent joint damage and a reduced quality of life.

The ability to predict response to first-line DMARDs, has been one of the main challenges in RA management for over two decades (3), and if solved would allow personalized therapy for the most severely affected patients. The added benefit of rapid control of inflammation would both reduce bone erosions and exposure to potentially toxic and ineffective drugs. Whilst factors such as gender, smoking and disease duration give a broad indication of likely treatment response, there are no specific markers (including autoantibodies and cytokine levels) that can accurately predict if DMARDs will control inflammation in RA (reviewed in(4)).

The role of B cells in the pathogenesis of RA

The efficacy of B cell depletion in patients with RA, (which can result in a profound and long lasting therapeutic effect in those that respond (5, 6)) has provided strong evidence that B cells are important effectors of RA. Rituximab works by targeting non-proliferating memory B cells and their immediate progeny, which includes the short-lived plasmablasts, that give rise to auto-antibodies. However, anti-citrullinated peptide antibodies (ACPA), which are specific for RA, can be present for more than a decade before clinical disease becomes apparent(7), and Atacicept treatment, which depletes antibody secreting plasma cells but not B cells is ineffective in multiple sclerosis (8) and rheumatoid arthritis(9). Together these observations make it unlikely that B cell depletion therapy works simply by depleting autoantibodies. Indeed as well as their role as antibody producers, B cells also secrete cytokines (including $\text{TNF}\alpha$, IL-10, IFN γ and IL6(10)), in combinations which vary with the stage of B cell differentiation and the inflammatory microenvironment [reviewed in(11)]. The secretion of IFN γ is particularly pertinent to RA as this cytokine activates macrophages to secrete $\text{TNF}\alpha$ and changes in macrophage markers have been shown to be a sensitive indicator of therapeutic response to rituximab therapy(12). In established RA, B cells likely become the main antigen presenting cells(13), moulding both the pathogenic T cell and macrophage pro-inflammatory responses [reviewed in(14)]. Hence B cells play a complex role in RA and we would hypothesise, based on the remarkable, novel findings discussed in more detail below, that they may drive DMARD refractory inflammation.

Currently one of the most important questions facing rheumatologists treating patients with RA is how to predict which patients will not respond to first line disease modifying drugs (DMARDs). Patients are only offered biologic therapy if they fail to respond to two conventional DMARDs. Currently, whilst it is appreciated that biologic therapy with cytokine blocking drugs is very effective, their high cost and potent immunosuppressive potential precludes their use in all patients. It takes at least a year to determine if patients will fail to respond to DMARDs, during which time further joint damage may have led to permanent disability. If it were possible to predict who would fail to respond to DMARDs and require biologic therapy at the time of initial diagnosis then patients could be reasonably offered biologic therapy earlier.

Statements of the Invention

According to a first aspect, the present invention provides a method of determining the likelihood of a subject responding to conventional DMARDs, the method comprising:

- a) determining the B cell hypermutation rate of said subject;
- b) determining therefrom the likelihood of the subject to respond to conventional DMARDs.

5 The invention is based upon the observation that patients who are unresponsive (refractory) to conventional DMARDs typically have lower hypermutation rates than those who are responsive. The reason for this difference is currently unclear. However, it provides a very useful indicator of the likely responsiveness of a patient to conventional first line treatments, and allows a clinician to make more informed decisions about the best treatment protocol for
10 a rheumatoid arthritis patient.

Hypermutation rate is suitably determined by:

- sequencing at least one region of interest of nucleic acids encoding the variable region of immunoglobulins (antibodies/B cell receptors) obtained from B cells from
15 said subject to obtain a set of sequence reads;
- comparing the determined sequence reads with corresponding germline allele sequences; and
- determining the number of nucleotide differences present in the sequence reads and thereby determining the hypermutation rate.

20 Hypermutation rate is preferably determined for IgG, Hypermutation rate for IgM can also be determined, but the inventors have found that IgM typically exhibits lower hypermutation rates than IgG, and that correlation of IgM hypermutation rate to response to DMARDs is not strong.

25 The method may comprise providing a suitable biological sample from the subject, the sample comprising B cells. Suitably the sample is a blood sample. The method may comprise obtaining a suitable biological sample from the patient, e.g. a taking blood sample.

30 Suitably peripheral blood mononuclear cells (PBMCs) can be obtained from a blood sample taken from the subject. B cells can be purified from PBMCs using well-known techniques; for example B cells can be purified on the basis of the presence of CD19, conveniently by magnetic-activated cell sorting (MACS) or the like. Alternatively, whole blood preparations or isolated PBMC preparations may be used in this assay.

Suitably the method comprises isolating mRNA or total RNA from B cells, preferably substantially the complete complement of mRNAs or total RNA. This can be readily achieved using well-known techniques. This mRNA or total RNA contains mRNAs encoding B cell immunoglobulins. From the mRNA or total RNA so obtained cDNA can suitably be synthesised using conventional reverse transcription technology. cDNA is more stable than mRNA and thus can be handled and processed more easily. Genomic DNA sequences from the B-cells may also be used as the source of nucleic acid.

Irrespective of the exact methodology used to obtain the nucleic acid sequences encoding immunoglobulins from the B cells, the method preferably involves amplifying at least a region of interest of the nucleic acids encoding immunoglobulin variable regions, e.g. by PCR using suitable primers flanking the region of interest. Suitably the primers for use in amplifying the immunoglobulin variable regions are directed to sequences within framework region 1 and the IgG and/or IgM constant regions. It will be apparent to the skilled person that there is freedom to select suitable primers depending on the exact region of interest to be amplified.

Suitably the complete variable region of the immunoglobulin heavy chain and/or immunoglobulin light chain is analysed; this is comprised of the region between leader sequence and constant region. However, it is possible that shorter regions within the above boundaries could be analysed.

Libraries of sequences comprising the region of interest are sequenced. There are several conventional sequencing technologies that can be used for this purpose. For example, an Illumina MiSeq instrument can be used, but there are a host of other options available to the skilled person.

The method may involve applying a quality control process to the sequence reads to eliminate any reads of dubious quality, e.g. those outside of an expected length range or with unexpected sequence properties. Reads may be discarded, for example, owing to low sequence quality, paired-end read merging disagreements or poor alignment to germline alleles. Optionally, where several sequence reads having the same sequence are obtained, the extra copies can be discounted.

Once the sequences of the region of interest have been determined, they are then compared with the corresponding germline allele sequence to determine the number of nucleotide differences present. This is typically performed for all relevant sequence reads.

Germline allele sequence can suitably be determined by statistical analysis of the sequenced variable regions. This avoids the need to determine the actual germline allele sequence experimentally. For example, the well-known VDJfasta utility can be used for this purpose, as discussed below.

5

It is generally preferred that the number of nucleotide differences present is determined across substantially all of the nucleic acids encoding the entire variable region of immunoglobulins from B cells, typically substantially the entire variable region of both the heavy and light chains; thus the region of interest can be substantially the entire variable region. For example, in the specific example described in detail below the entire variable region up to the junction with CDR3 was used – in other words the region which, reading in a sense direction, begins at the end of the immunoglobulin leader sequence and ends at the beginning of the constant region.

15 However, it is perfectly possible that the number of nucleotide differences present could be determined for at least one less extensive region of interest, e.g. in only the heavy chain or the light chain, across one or more CDRs in one, other or both of the heavy and light chains, including or excluding framework regions, CDR3 alone, etc. Hypermutation is most prevalent in the CDRs, and thus focussing on CDR sequences as regions of interests may
20 be a suitable approach in some cases.

A region of interest can be obtained through the appropriate selection of primers so that the desired region is amplified, or it can be obtained using conventional bioinformatics techniques to select the region of interest from sequence data for a larger region.

25

In preferred embodiments of the present invention, for each sequence read obtained, the number of nucleotide differences in the region of interest between the sequence read and the corresponding germline V-region allele is determined. This is readily achieved using appropriate software (such as the VDJfasta utility discussed herein). A region of alignment
30 between the sequence read and the germline allele can be defined to allow the comparison to be easily made.

Once the number of sequence differences in the region of interest has been determined, a hypermutation rate can be determined. The hypermutation rate should be determined in a manner which allows comparison between the observed hypermutation rate of a subject and
35 'normal' hypermutation rate, i.e. that observed in healthy controls or, more preferably,

responsive RA patients. The hypermutation rate can be defined in many ways, and some non-limiting examples are given below.

In some embodiments of the invention the hypermutation rate can suitably be defined as a number of nucleotide differences observed per a given sequence length, e.g. X changes observed per 100 bp, 200 bp or the like. This can readily be determined in view of the total number of sequence changes observed and the total length of all sequences analysed. Such a hypermutation rate is thus a simple representation of the amount of hypermutation which has occurred in a subject, a lower value being indicative of a lower levels of hypermutation.

Alternatively, the hypermutation rate can suitably be defined as the mean number of hypermutations per sequence read. This is the total number of nucleotide differences observed in all reads (or in a defined portion of all reads) divided by the number of sequence reads analysed.

In another embodiment the likelihood of a patient being a DMARD non-responder may be determined with reference to a threshold defined as the proportion (e.g. percentage) of sequence reads which demonstrate a given occurrence level of nucleotide differences, typically an atypically low occurrence of nucleotide differences. A subject having a low hypermutation rate would typically have a higher proportion of sequence reads having a low number of nucleotide differences when compared to a reference value (e.g. from a control population or a population of responders). Specific values are impossible to define, because they will be determined by the specifics of the region(s) of interest analysed. By way of example, in the specific results described below, the percentage of reads exhibiting fewer than 5 nucleotide changes revealed a dramatic difference between non-responders and responders. For a shorter region of interest than that described below, the proportion of sequences showing fewer than 4, fewer than 3, fewer than 2, or fewer than 1 per read could be used.

In step b) of the method, the determination may comprise comparing the determined B cell hypermutation rate with a reference value. For example, where the determined B cell hypermutation rate is lower than the reference value, the subject can be identified as a likely non-responder to conventional DMARDs. Typically the most important consideration for a reference value is that it is indicative of a significant level of reduction in B cell hypermutation rate compared with a 'normal value', e.g. healthy controls or RA responders.

Such a reference value can suitably be derived from a reference B cell hypermutation rate which has been determined from healthy controls, or, more preferably, from RA patients who are responsive to conventional DMARDs.

- 5 The reference value can be a value selected on the basis of defining a statistically significant threshold.

More than one reference value can be used, with each reference value representing a different level of significance or clinical implication.

10

Wherein in step b) the hypermutation rate is determined across the entire variable region, a non-responder may be indicated by a mean number of hypermutations per sequence read of 17 or fewer, for example 16 or fewer, 15 or fewer, 14 or fewer, 13 or fewer, 12 or fewer, 11 or fewer, or 10 or fewer. As discussed below, responders typically have a mean hypermutation rate of 18.69, and controls have a mean hypermutation rate of 20.69, compared with 10.92 for non-responders. Accordingly a value of less than 18.69 is indicative of non-responsiveness, with successively lower values representing a stronger indication. However, it will be appreciated that the skilled person can select an appropriate reference value depending on the details of any given method.

15

20

Alternatively, the method may comprise generating a graphical representation of the hypermutation rate. For example a graph of the number of nucleotide changes per sequence read versus the number of reads displaying that number of changes. The shape of such a graph provides a clear indication of changes to hypermutation rate of a subject.

25

Alternatively, the method may comprise the comparison of the distribution of hypermutation rates for all or selected sequences derived from a patient with that generated from a reference sample. This comparison may be made using statistical tests or machine-learning algorithms. For example, distributions may be compared using a skewness function, as illustrated in the results described below.

30

The method can comprise the step of selecting, based upon step b), a subject for immediate or early treatment with biologic DMARDs. By immediate is meant without attempting treatment with a conventional DMARD prior to administration of a biologic DMARD, and by early is meant that treatment with a conventional DMARD(s) would be tried for a shorter period than conventional.

35

Thus the method may be used by a clinician to assist in deciding upon the most appropriate treatment for an RA patient. For example, whether the patient should initially be placed on conventional DMARDs at all (e.g. using biologics from the outset for predicted non-responders), whether the patient should be treated with a conventional DMARD initially but
5 with response to treatment being reviewed early, or whether the patient should be treated in a conventional manner (e.g. typically administering at least two conventional DMARDs for a period of 12 months to assess responsiveness).

In some embodiments the method is applied to subject diagnosed or suspected of having an
10 inflammatory arthritis. The inflammatory arthritis is suitably RA, but it can be another form of inflammatory arthritis, such as psoriatic arthritis (PSA) and/or other spondyloarthropathies. Where the inflammatory arthritis is RA, the RA can be seropositive or seronegative. In some preferred embodiments the RA is seropositive RA, for example early, untreated seropositive
15 RA.

In a second aspect the present provides a method of treating a subject having an inflammatory arthritis (suitably RA, e.g. seropositive and/or seronegative RA), the method comprising the steps of:

- a) providing a biological sample from a subject comprising B cells;
- 20 b) determining the B cell hypermutation rate of said subject;
- c) determining therefrom the likelihood of a patient to respond to conventional DMARDs; and
- d) administering an appropriate treatment to the subject based at least partially upon the determination in step c).

25 For example, where the subject has a B cell hypermutation rate below a predetermined reference value the subject may be treated immediately or early with a biological DMARD. For example, a biological anti-TNF α treatment may be administered, e.g. an anti-TNF α antibody or decoy TNF α -receptor.

30 According to a third aspect the present invention provides an assay system for determining the likelihood of a subject (e.g. a subject suffering from an inflammatory arthritis, e.g. RA) to respond to non-biologic DMARDs, the assay system being adapted to carry out the method as defined in the first aspect of the invention.

35 Suitably the assay system comprises:

- one or more reaction vessels to receive a biological sample comprising B cells of said subject;
- nucleic acid isolation means to isolate nucleic acids which encode the variable region of immunoglobulins (antibodies/B cell receptors) from said B cells of said subject;
- 5 - nucleic acid amplifying means adapted to amplify at least one region of interest of the nucleic acids within the sequences encoding the variable region of immunoglobulins;
- sequencing means to determine the sequences of the amplified nucleic acids including the at least one region of interest; and
- computing means configured to run suitable software to determine germline allele
10 sequences and determine the number of nucleotide changes in the sequences obtained for the region of interest compared to the corresponding germline allele sequences.

The nucleic acid amplifying means suitably comprises suitable primers to amplify the nucleic acids including the region of interest and suitable reagents to perform PCR. Suitable
15 primers are set out below, but various other primers could be used, and could be readily determined by the person skilled in the art.

In some cases direct nucleotide sequencing, particularly direct RNA sequencing, without the amplification can be used, in which case nucleic acid amplification means can be omitted.
20 For example, suitable sequencing approaches are described in Ozsolak *et al.* 'Direct RNA sequencing', Nature 461, 814-818 (8 October 2009), and RNA sequencing services are available commercially from companies such as SeqLL, LLC (Woburn, MA 01801, US).

In a fourth aspect the present invention provides a kit of parts, the kit of parts for use in a a
25 method as set out above, the kit comprising suitable primers to amplify nucleic acids including the region of interest, and optionally suitable reagents to perform PCR and/or suitable probes for performing quantitative PCR, e.g. real time PCR. Suitable primers are set out below, but various other primers could be used, and could be readily determined by the person skilled in the art.

30 To facilitate the understanding of this invention, a number of terms are defined or explained below. Terms defined herein have meanings as commonly understood by a person of ordinary skill in the areas relevant to the present invention. Terms such as "a", "an" and "the" are not intended to refer to only a singular entity, but include the general class of which a
35 specific example may be used for illustration. The terminology herein is used to describe specific embodiments of the invention, but their usage does not delimit the invention, except as outlined in the claims.

'Hypermutation' (or somatic hypermutation (SHM)) is a cellular mechanism by which the immune system adapts to the new foreign elements that confront it. Somatic hypermutation is part of the affinity maturation process and involves a programmed process of mutation affecting the variable regions of immunoglobulin genes. Unlike germline mutation, SHM affects only individual immune cells, and the mutations are not transmitted to offspring.

'Hypermutation rate' is an objective measure of the amount of hypermutation which has occurred in a subject. It can be determined by comparing the observed variable region sequences and the original germline allele sequence. The number of nucleotide differences which are observed compared with the germline sequences can simply be counted. The hypermutation rate can then be defined in an appropriate manner, e.g. as discussed above. Typically, hypermutation rate refers to the degree of sequence change within the B-cell receptor sequence repertoire of a subject, which arise through mutations, such as during the somatic hypermutation process.

"Conventional DMARDs" or "non-biologic DMARDs", or variations thereof, as used in the in the context of the present invention relate to non-protein therapeutic agents, in particular DMARDs which are not antibodies or cytokine receptors. For example, conventional DMARDs include methotrexate (MTX), hydroxychloroquine (HCQ), sulfasalazine (SSZ), and leflunomide, which are also typical "first line" conventional DMARDs.

"Biologic therapies" or "biologic DMARDs", or variations thereof, as used in the context of the present invention relates to the well-known family protein-based RA (and other inflammatory arthritis) therapeutic agents. Typically these therapeutic agents comprise antibodies or modified (decoy) receptor proteins which are adapted to modulate with the activities of immunomodulatory agents or immune processes. For example, the term commonly relates to anti-TNF α antibodies or modified (decoy) TNF-receptor proteins which interfere with the activity of the pro-inflammatory cytokine TNF- α . Exemplary biologic therapies include adalimumab (anti-TNF α antibody), etanercept (decoy TNF α receptor), golimumab (anti-TNF α antibody), certolizumab (PEGylated Fab' anti-TNF α), infliximab (anti-TNF α antibody), abatacept (T-cell co-stimulatory signal inhibitor), tocilizumab (anti-IL6 antibody) and rituximab (anti CD20 and B cell depleting).

Embodiments of the present invention will now be described, by way of non-limiting example, with reference to the accompanying drawings.

Brief Description of the Figures

Figure 1 Schematic diagram of B cell repertoire sequencing process. PBMCs were derived from blood samples taken from individuals with RA or healthy donors and B cells were purified by CD19 MACS. Following mRNA purification, cDNA was synthesized and immunoglobulin variable regions were amplified by PCR using primers within framework region 1 and the IgG and IgM constant regions. Variable region amplicon libraries were sequenced using an Illumina MiSeq instrument.

Figure 2 V/J allele usage and diversity of antibody repertoires Panel A: Dotplot of VJ usage combinations for all samples. Plots show the percentage representation of each Valleele-Jallele combination for all samples. Panel B: Diversity of antibody repertoires. D50 values for each donor were calculated as the minimum percentage of CDR3 clones that can account for 50% of the total sequence repertoire sampled. Comparison of the group means by non-parametric test did not indicate a significant difference between group mean diversity scores ($p=0.51$, Mann-Whitney test).

Figure 3 Mean length of CDR3 from individuals with RA or healthy donors. The CDR3 length of each sequence was determined by prediction of the CDR3 region using the VDJfasta Ig alignment tool. Points in the dotplot show the mean CDR3 length for each individual in either the rheumatoid arthritis (RA) or healthy donor (HD) groups. There was no significant difference between the mean CDR3 lengths of each group ($p=0.26$, t-test).

Figure 4 Hypermutation rates of donor IgG sequences. For each antibody sequence, the closest known germline V-segment allele was determined. The number of mutations relative to the closest predicted germline allele was calculated and plotted. Panel A shows density plots of the IgG hypermutation rates of the mean number of hypermutations within the V-segment for sequences of IgG isotype from each donor from either the healthy donor, rheumatoid arthritis DMARD responder or 1° DMARD non-responder (1°DNR) groups. Panel B is a graph of the skewness of the IgG hypermutation distribution for each sample. The measured skewness is greater in samples from the RA group compared to the control group (t-test with Welch's correction, $p=0.038$). Panel C shows the mean IgG hypermutation rate for each sample, grouped by patient outcome.

Figure 5: Heavily skewed IgG hypermutation rates are associated with failure to respond to DMARDs. Panel A: Cut-offs on IgG Hypermutation data. Density plot of all reads from healthy donors (black), DMARD Good Responders (dot-dash line) and 1°-DNRs (dashed line) from B cells sampled at diagnosis. In the 1°-DNRs the distribution of IgG hypermutation rates is heavily skewed towards zero, whereas GR RA patients have distributions with only a slight negative skew. The percentage of reads with fewer than five mutations was calculated for each donor, and this threshold is indicated on the chart by a vertical dotted line. Panel B shows percentage of IgG reads below threshold of 5 mutations for each donor. IgG

repertoires from RA donors at diagnosis or when referred to biologics clinics. Donor groups are split according to whether they responded initially to DMARDs (GR) or were found on follow up to be refractory to DMARDs (1^o-DNRs). IgG repertoires from 1^o-DNRs had highly skewed hypermutation rates, leading to higher percentages of reads with fewer than 5 mutations (Statistics-Mann-Whitney non-parametric test).

Figure 6: This shows the skewness for each sample, split by IGHV germline family. Skewness is consistently higher across all germline families, indicating a widespread B cell phenotype not restricted to specific germline families.

Figure 7: This shows a schematic of the primer system used to amplify the region of interest (it is shown in a simplified form because there are many barcoded variants of the constant region primers). Fig 7a shows the primer structure and their location relative to the IgHV region. Fig 7b shows exemplary primer sequences used in the specific experiments described below. The gene-specific portions of the primers can readily be replaced, and any suitable gene specific primer for the V-region or leader sequence could be used. As discussed below, suitable primer systems are described in detail in the prior art.

Figure 8: Hypermutation frequencies of IgG Vh sequences from DMARD patients in biologics cohort. Panel A: Individual density plots of the IgG mutation frequencies of 47 RA donors attending biologics clinic. Primary non-responders (1DNR-RA) are filled with diagonal lines and other biologics patients (DMARD^R-RA) are filled with a hashed pattern. Y axes are normalized by scaling probability density function values for each graph to the range 0 to 1. Panel B: Percentage of IgG reads below threshold of 5 mutations for each donor in the biologics cohort. RA donors that were refractory to DMARD treatment (1DNR-RAs) had a higher percentage of sequences containing fewer than 5 mutations than patients who had responded to DMARDs (DMARD^R-RA) ($p < 0.001$, Mann-Whitney test). Horizontal bars denote median values.

Figure 9: Ages of donors from the biologics cohort. Horizontal lines indicate mean age for each group.

Figure 10: Mutation count distributions from samples taken in Tempus blood RNA tubes. IgG repertoires were sequenced from 3 mL blood samples extracted into Tempus RNA tubes and stored at room temperature for 3 days. Density plots show the distribution of the number of mutations per sequence read for the repertoires of 7 RA patients and 3 healthy donors (HD).

Specific Description of Embodiments of the Invention

While embodiments of the present invention are discussed in detail below, it should be appreciated that the present invention provides many applicable inventive concepts that can

be embodied in a wide variety of specific contexts. The specific embodiments discussed herein are merely illustrative of specific ways to make and use the invention and do not delimit the scope of the invention.

5 The central hypothesis of the present project was to determine if patients with rheumatoid arthritis, expressed a signature of specific B cell receptors (BCRs), which could be elucidated following massively parallel B cell sequencing. Of particular interest was early, untreated seropositive rheumatoid arthritis where failure to provide effective treatment at an early stage can result in significant joint damage. However, seronegative RA and other
10 inflammatory arthritis conditions such as PSA and other spondyloarthropathies were also of interest, as conventional DMARDs are used in these conditions with variable efficacy levels. The data reported herein describes the B cell signature of patients with a new diagnosis of rheumatoid arthritis. In addition it identifies a specific set of BCRs that identify those patients that will fail to respond to conventional DMARD therapy and which will require biologic
15 therapy at a later date, possibly after significant joint damage has occurred. It opens the way for further studies into the pathogenesis of the disease and to a means to target a specific pathogenic B cell subset.

Materials and Methods

20

Ethical review and donor selection criteria

This study was conducted in accordance with the guidelines of the World Medical Association Declaration of Helsinki (15), and the collection and use of human samples was approved by the South East Scotland Bioresource NHS Ethical review board. Informed
25 consent was obtained from all study participants prior to sample collection.

The following inclusion criteria were applied to RA donors for inclusion in this study: donors were diagnosed with Rheumatoid Arthritis by a Consultant Rheumatologist, tested for both ACPA and rheumatoid factor (sero-positives) and had not previously received any treatment
30 for rheumatoid arthritis. Control donors had no current or past history of rheumatoid arthritis and were specifically selected to obtain a donor age range that reflected the RA cohort as closely as was possible.

Cell purification and B cell repertoire sequencing

35

Blood samples were collected buffered in tri-sodium citrate and immediately stored at 4°C to be processed within a few hours. Peripheral blood mononuclear cells (PBMCs) were isolated

by ficollpaque density centrifugation using Leucosep tubes (Greiner Bio-One). Blood was diluted 2-fold in RPMI medium (Sigma Aldrich, UK), transferred to pre-filled Leucosep tubes and centrifuged for 15 minutes at 800 x G. Most of the solution above the interface was removed using a pipette and the solution around the interface was transferred to a new tube.

- 5 Cells were washed three times in RPMI + 10% FCS then frozen in aliquots of 10^7 cells per aliquot in 10% DMSO, 90% FCS.

Individual aliquots of PBMCs from each donor were thawed and B cells were purified by CD19 MACS according to the manufacturer's recommended protocol. Messenger RNA (mRNA) was purified from 10^6 B cells using oligo-dT microbeads according the manufacturer's instructions and on-bead first strand cDNA synthesis was performed using Superscript III first-strand synthesis supermix (Invitrogen). V-region amplicons were generated by PCR using individual pools of forward primers within framework region 1 (FR1) that were designed to amplify all known V-region alleles, and a reverse primer within the IgG and IgM constant regions (see Figure 7). Both primer sets incorporated Illumina P5 or P7 adaptor sequences at their 5' ends to facilitate sequencing (sequences provided below). The sample preparation process is shown in Figure 1. Amplicons were purified using an eGel Size-Select electrophoresis system (Life Technologies, UK) to select products within the anticipated size range of approximately 400-450bp. 250bp paired-end sequencing was performed on an Illumina MiSeq sequencer using a pool of read1 sequencing primers matching the pool of FR1 primers but omitting the Illumina adaptor sequence, an indexing primer to provide indexing information and two read2 primers matching the IgG and IgM constant region amplification primers that also lacked the Illumina adaptor sequence.

- 25 P5 adaptor – AATGATACGGCGACCAACGAGATCTACAC (SEQ ID NO 1)
P7 adaptor – CAAGCAGAAGACGGCATACGAGAT (SEQ ID NO 2)

IGHV FR1 forward amplification primers (Format: P5 Illumina adaptor | PAD Sequence | Gene-specific sequence):

- 30 Name: hIGHV2.0 (SEQ ID NO 3)
AATGATACGGCGACCAACGAGATCTACAC|GACATGACTTCGT|CCCTCTCACTCACCTGTG
Name: hIGHV2.1 (SEQ ID NO 4)
AATGATACGGCGACCAACGAGATCTACAC|GACATGACTTCGT|CCCTKAGACTCTCCTGTG
Name: hIGHV2.2 (SEQ ID NO 5)
35 AATGATACGGCGACCAACGAGATCTACAC|GACATGACTTCGT|CCCTGARACTCTCCTGTG
Name: hIGHV3.0 (SEQ ID NO 6)
AATGATACGGCGACCAACGAGATCTACAC|GACACTACAGCGT|AGACCCTCACRGTGACCT

Name: hIGHV0.0 (SEQ ID NO 7)

AATGATACGGCGACCACCGAGATCTACAC|GACATAGACCGGT|TCAGTGAAGGTYTCCTGC

Name: hIGHV1.0 (SEQ ID NO 8)

AATGATACGGCGACCACCGAGATCTACAC|CGCCGGAAGCAGT|GCTGAGGTGAAGAAGCCT

5 Name: hIGHV1.1 (SEQ ID NO 9)

AATGATACGGCGACCACCGAGATCTACAC|CGCCGGAAGCAGT|YCAGGACTGGTGAAGCCT

Name: hIGHV1.2 (SEQ ID NO 10)

AATGATACGGCGACCACCGAGATCTACAC|CGCCGGAAGCAGT|SCAGGACTGTTGAAGCCT

Name: hIGHV4.0 (SEQ ID NO 11)

10 AATGATACGGCGACCACCGAGATCTACAC|CATGTGCCTGTGT|GTCCCTGAGACTCTCCTG

Name: hIGHV4.1 (SEQ ID NO 12)

AATGATACGGCGACCACCGAGATCTACAC|CATGTGCCTGTGT|GTCTCTGARGATCTCCTG

Name: hIGHV5.0 (SEQ ID NO 13)

AATGATACGGCGACCACCGAGATCTACAC|CTTAGAGTCACGT|CTCCTGCAAGGYTTCTGG

15

IgGHC reverse primer (Format: P7 Illumina adapter | Barcode | PAD | Gene-specific sequence):

Name: hIGHHVrev(INDEX1) (SEQ ID NO 14)

CAAGCAGAAGACGGCATACGAGAT|AGCTCT|AGTCAGTCAGCC|ATGGGCCCTTGGTGG*A

20

IgMHC reverse primer (Format: P7 Illumina adapter | Barcode | PAD | Gene-specific sequence):

Name: hIGMHVrev(INDEX1) (SEQ ID NO 15)

CAAGCAGAAGACGGCATACGAGAT|AGCTCT|AGTCAGTCAGCC|GGCGGATGCACTCC*C

25

Alternative primers to amplify all IGHV regions are widely available and publicly disclosed (e.g. those listed in Schanz M, et al. (2014) High-Throughput Sequencing of Human Immunoglobulin Variable Regions with Subtype Identification. PLoS ONE 9(11): e111726. doi:10.1371/journal.pone.0111726, or available at the well-known IMGT immunoglobulin database <http://www.imgt.org/IMGTPRIMERDB/>). In addition, various groups have also amplified B-cell variable regions without using V-region specific primers, but instead using a 5'RACE approach using the SMART system (see e.g. He et al., 2014, <http://www.ncbi.nlm.nih.gov/pubmed/25345460>).

30

35 **Sequence Quality Control, paired end joining and filtering**

Sequence read-pairs were combined using the Flash utility (16). Sequence pairs which did not meet the quality criteria or which were shorter than 300bp once combined were excluded from further analysis.

Sequence analysis

All joined sequences were processed using the VDJfasta utility (17). VDJFasta uses a Hidden Markov Model to statistically analyse sequences upstream and downstream of putative CDR3s and outputs V, D and J germline sequences, CDR3 sequences and translated protein sequences derived from each sequencing read. Sequence processing using this utility was parallelized using custom Perl and Python scripts and run on the parallel computing facility provided by the Edinburgh Compute and Data Facility (ECDF, <http://www.ecdf.ed.ac.uk/>). Sequence isotope was predicted according to the constant region sequence present directly upstream of the reverse sequencing primer.

CDR3 length

CDR3 sequences were extracted for each sequence and the mean CDR3 sequence length for each donor was calculated inclusive of the boundary cysteine and tryptophan residues. Mean CDR3 hydrophobicity for each CDR3 was calculated as the GRAVY score for the component residues using functions from Biopython libraries ([http:// www.biopython.org/](http://www.biopython.org/)).

Hypermutation rate

For each sequence read, the number of nucleotide differences in the V-region between the sequencing read and the predicted germline V-region allele across the region of alignment was calculated using the VDJfasta utility. Data were aggregated by donor and the mean number of mutations within the V-region was plotted for each donor. Frequency distributions for each donor were derived using the R statistical package using the 'ggplot2' and 'plyr' packages. Skewness calculations were performed using the skewness function of the R package 'moments' (18) which calculates the degree of asymmetry within the hypermutation data distribution (19).

Diversity calculation

Repertoire diversity was assessed by calculation of D50 values for each donor by Perl implementation of the formula previously described (20). The D50 is defined as the minimum percentage of unique CDR3 clones that can account for 50% of the total number of sequencing reads.

Single linkage clustering

All CDR3 sequences were extracted from the translated Vh amino acid sequences and de-replicated using USearch 6 64-bit version (21), maintaining counts for individual sequences. A sequence identity matrix for all possible pairwise comparisons of CDR3 sequences was

produced by sequence alignment and scoring using the Smith-Waterman algorithm (22). From this sequence matrix, single-linkage clustering was performed with lower identity threshold of 90% to produce network graph file in GEXF format. This file was viewed and manipulated using Gephi toolkit (<https://gephi.org/toolkit/>). All analyses were performed, using parallelized code where possible, using the ECDF parallel computing facility at the University of Edinburgh.

Results

Sequencing of immunoglobulin variable region repertoires in patients with rheumatoid arthritis

Antibody variable region genes from rheumatoid arthritis patients and healthy controls were amplified from cDNA generated from CD19+ B cells purified from blood samples (see Figure 1). Sequencing of pooled amplicons from all patients generated a dataset of 11,295,381 sequence reads (mean number of sequences per donor 364,367, minimum 154,931, maximum 595,687). Following paired-end read merging and processing of sequences with the VDJfasta utility, a total of 10,827,709 sequences were retained in the dataset and the remainder discarded owing to low sequence quality, paired-end read merging disagreements or poor alignment to germline alleles.

VDJ usage and diversity of antibody repertoires

To investigate whether there was any bias in frequencies of particular VDJ recombinants between groups, germline V, D and J alleles were predicted for each sequence. Owing to the lack of confidence surrounding D allele prediction resulting from junctional nucleotide additional or deletion during VDJ recombination, subsequent analyses used only the V and J allele calls for each sequence. The frequency of each combination of V and J alleles was calculated as a percentage of the total number of reads for each sample. The percentage of sequences that were predicted to originate from each germline V and J allele combination is shown as a dot plot format in Figure 2 panel A.

The percentage incidence of each V/J germline allele combination was compared between the donors in the RA group and the healthy donors using a Kruskal-Wallis non-parametric test and Dunn's post-test with Bonferroni correction for multiple comparisons. None of the p-values were lower than the FDR-corrected alpha value of 0.0001006, indicating that no V/J combinations were significantly altered in the RA group as compared to the healthy control group.

We investigated the hypothesis that there may be a difference in repertoire diversity in RA patients as a result of the expansion of pathogenic B cell populations in disease. Using the D50 measurement of repertoire diversity suggested by Han *et al* (US20120183969/WO2012/097374) for TCR repertoires, we calculated the D50 values for each sample. The range of D50 values were found to be wide in both the RA donors (median 1.86, range 0.035 to 13.960) and the control donors (median 0.94, range 0.096 to 12.070). The mean D50 values were not found to significantly different between the groups ($p=0.51$, Mann-Whitney test). D50 values are shown in Figure 2, panel B.

Analysis of CDR3 properties

The mean CDR3 sequence length for each sample is plotted in Figure 3. Mean CDR3 length in the RA group ranged from 15.6 to 18.5 amino acids and in the control group ranged from 16.1 to 17.4 amino acids. No inter-group difference in mean CDR3 length was detected by statistical analysis (t-test, $p=0.26$).

Hypermutation rates

For each sequence, we compared the V-segment read sequence with the predicted germline allele sequence to determine the hypermutation rate, which was calculated as the number of nucleotide changes within the sequenced region of the V-segment, up to the junction with CDR3 (which can be termed the 'region of interest'). Density plots of the number of V-region hypermutations per IgG sequence for each sample are shown in Figure 4a. A skewing of the hypermutation distribution towards lower hypermutation rates was observed in a number of the rheumatoid arthritis samples but not in the healthy donor samples. The magnitude of this skew, or 'skewness', was calculated from the hypermutation frequency distribution for each sample (Figure 4b). The mean skewness of hypermutation rate distributions from the rheumatoid arthritis donors was significantly higher than the control donors (t-test, $p=0.0256$), indicating a skew towards a lower number of V-region hypermutations in the rheumatoid patients.

This negative skew in IgG hypermutation rate was particularly pronounced in three of the RA patient samples: A0003, A0010 and A0011. Examination of clinical records revealed that, remarkably, all three of those RA donors failed to respond adequately to first-line DMARD treatment at a time point of 12 months since blood sampling at initial clinical consultation. All other RA patients who participated in the study were assessed as responding well to DMARD treatment at this point. Figure 4c shows the mean hypermutation rates when samples were split according to clinical outcome. The mean number of hypermutations per sequence was significantly lower in the non-responding patients than the responding

patients or the control donors (group mean 10.92 vs 18.69 or 20.69, ANOVA with Tukey's range test). Figure 5a compares normalised density versus number of mutations for all three groups, clearly demonstrating the lower levels of hypermutation in non-responders and Figure 5b further illustrates that percentage of reads having fewer than 5 mutations was far higher in non-responders.

Discussion

B cells play a fundamental role in the pathogenesis of RA, which is seen in the effectiveness of B cell depletion therapy using anti-CD20 monoclonal antibodies in severe disease. The mechanisms by which B cells are suggested to contribute to the self-perpetuating cycle of auto-immune inflammatory responses in the synovium include the production of rheumatoid factor and ACPA, contribution to T-cell activation by acting as antigen-presenting cells, and the production of pro-inflammatory cytokines. This led us to study the peripheral blood B cell repertoire in RA and healthy control donors by massively parallel sequencing to look for differential characteristics that might indicate the presence of pathogenic clonotypes or alterations to the B cell repertoire. We did not find the overall repertoire to be altered at the level of V and J combinatorial frequencies, as might be expected if conserved clonotypes of B cells were responding to disease. Indeed, we did not find that any of the conserved CDR3 clonotypes were more highly represented in the RA repertoires. CDR3 properties were not differential between health and disease, confirming that there were no global alterations to CDR3 characteristics. However, distinctions were observed in the hypermutation rates of repertoires from RA and healthy donors. This was indicated by a negative skew in the hypermutation distributions in a number of the RA repertoires, which was particularly distinct only in the cases of the three patients who failed to respond to DMARDs within the twelve month period of the study. This finding leads us to postulate that it should be possible to use B cell repertoire analysis to predict the patient's future response to DMARDs and thereby commence biologic therapy earlier than is presently possible.

The availability of such a test would address one of the most important questions facing rheumatologists treating patients with RA: how to predict which patients will not respond to DMARDs and should therefore be progressed to biologic therapy at the time of diagnosis. Currently patients are only offered biologic therapy if they fail to respond to two DMARDs, (given for a sufficient length of time to determine inefficacy or intolerance), such as methotrexate, sulphasalazine or hydroxychloroquine. Currently, whilst it is appreciated that biologic therapy with cytokine blocking drugs is very effective, their high cost and potent immunosuppressive potential precludes their use in all patients. It takes at up to a year to determine if patients will fail to respond to DMARDs, during which time further joint damage

may have led to permanent disability. Further expansion of this study to increased patient numbers will be necessary to confirm, *a priori* rather than *post hoc*, that this finding is reflective of the wider population of inflammatory arthritis patients. However, the present findings are very promising and highly statistically significant. The present invention thus facilitates the deployment of a convenient, simple and reliable test that can be used to identify DMARD resistant patients at the time of diagnosis.

Acknowledgements

This work has made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (<http://www.ecdf.ed.ac.uk/>).

Further Experimental Work

1) Hypermutation rates in a further, independent "biologics cohort"

The inventors investigated whether the same IgG mutation frequency as noted above would be seen in a larger cohort of 47 seropositive RA patients who had already commenced DMARDs – referred to as the "biologics cohort". In this study all the RA patients had progressed to biologic DMARDs, due to a sustained high disease activity (DAS28>5.1 on at least 2 occasions and for at least 2 months), despite treatment with synthetic DMARDs. 15 out of 47 subjects were noted to have been 1°DNR-RA patients (Table 1, below). Importantly all of these donors also exhibited a dramatically skewed IgG mutation profile (shown with diagonal fill pattern in Figure 8A), with a significantly higher percentage of IgG V_H sequences containing fewer than 5 mutations (Figure 8B). These 1DNR-RA patients possessed IgG repertoires, in all of which over 18.5% of the total IgG V_H sequences had fewer than 5 mutations, with a median of 26.67%. In contrast, the remaining 30 patients sequences contained a median of 6.60% sequences containing fewer than 5 mutations, whose IgG V_H had all responded to DMARD therapy initially, but DMARD treatment had subsequently become secondarily ineffective or was poorly tolerated (Table 1). Based on the cohort tested, classification of patients with this threshold gave a test sensitivity of 100% [95% C.I. 78.2% to 100%] and a specificity of 100% [95% C.I. 89.1 to 100%].

This analysis confirms the initial findings that hypermutation rates provide a useful indicator of the likelihood of a patient responding to conventional DMARDs.

The materials and methods used in the above further hypermutation analysis work were essentially identical to the methods described above. However, whereas the methodology described above used mRNA, the further work on the biologics cohort used total RNA. In this

case total RNA was extracted using Trizol reagent and purified using a Directzol RNA purification kit (Zymo Research) prior to first strand cDNA synthesis (using Superscript III first-strand synthesis supermix (Invitrogen), as for mRNA).

5 2) Additional investigations

The inventors considered whether the skewing of V_H mutational profiles in the 1°DNR-RA patients resulted from large clonal expansions of B cells, possibly as a result of an auto-antigen driven response. However, an analysis of the diversity of the expressed V_H repertoire did not confirm this. Firstly, the low mutation frequency in the IgG sequences of 10 1° DNR-RA patients was not restricted to just one V_H family but was apparent in all V_H families. Secondly, sequences from the IgG V_H repertoire that exhibited low mutation frequencies were found in a large diversity of clones. Furthermore, looking at the most abundant CDR3s revealed no difference in the frequency of dominant clones as a percentage of the Ig repertoire. Finally, the diversity of the IgG V_H repertoires (as measured 15 by the D50 values¹⁵), calculated as the minimum percentage of unique CDR3 clones that can account for 50% of the total repertoire) did not differ between the two groups.

Another explanation for the low mutation frequencies in the 1°DNR-RA patient group might 20 be the prevalence, in the peripheral blood, of B cell populations that naturally exhibit few V_H mutations, for instance class-switched marginal zone B cells¹⁶. We analyzed the B cell subsets by multi-parameter flow cytometry. Although there were some noticeable trends, none of the B cell subsets showed any significant differences between the two patient groups. We found a significantly lower number of CD73⁺ IgG memory B cells in 1°DNR-RA 25 patients compared to healthy donors. However, this was also true for the DMARD responder group, indicating that it was not related to the lack of response to DMARD therapy. It is worth noting, in particular, that there is no increase in the proportion of plasmablasts or plasma cells in the circulation of 1°DNR-RA patients, compared to healthy controls. These cells, expressing large quantities of mRNA, could skew the repertoire analysis; the fact that they 30 are not expanded makes this less likely. In addition, if plasmablasts were the source of poorly mutated IgG V_H sequences, we would have expected to see hypomutated sequences restricted to large nodes within a network analysis, and this was not the case. The ease of analyzing IgG regions by next generation sequencing makes the translation of this observation highly feasible for patients with RA.

35

In the additional experimental work it was also noted that B-cell repertoire sequencing could also be readily performed on just 3 ml of peripheral blood drawn directly into blood RNA

vacuum tubes, even following storage for 3 days at room temperature (Figure 10) and the bioinformatics pipeline can be simplified by focussing testing on a smaller number of common V_H gene families.

5 The above experimental work was performed as follows:

CDR3 Clonal analysis of B-cell repertoires - The CDR3 sequences were de-replicated and a sparse distance matrix of sequence identities above 90% was produced by Smith-Waterman alignment of all combinations of CDR3 sequences. Clonal composition diagrams were produced using the OpenOrd force-directional layout algorithm of Gephi 0.8.

10

Flow Cytometry

Flow cytometry analysis of PBMCs from RA patients attending a biologics clinic was compared with those from healthy donors. The prevalence of B-cells populations as a percentage of the total B-cells were calculated by gating using fluorescently conjugated antibodies. Single cell suspensions of human PBMCs were fluorescently-labelled with antibodies to cellular markers according to manufacturer's instructions and cells were acquired on a LSR Fortessa Instrument (BD Biosciences): CD21-BV421(B-ly4), CD38-BV510(HIT2), CD73-BV605(AD2), CD19-PeCy5(HIB19), CD27-PC(M-T271), CD20-AF700(2H7), IgD-APC-H7(IA6-2) from BD Biosciences; IgM-BV650(MHM-88), CD24-FITC(ML5), CD138-PE-Cy7(MI15) from BioLegend; CD1c-PerCP-eF710(L161) from eBioscience and IgG-PE(IS11-3B2.2.3) from Miltenyi.

20

Table 1 - All the RA patients had progressed requiring biologic DMARDs to control severe RA, due to a sustained high disease activity (DAS28>5.1 on at least 2 occasions and for at least 2 months). All of the patients had shown a response to synthetic DMARDs (such as methotrexate, sulphasalazine, leflunomide or hydroxychloroquine).

25

Sub.	Age	Sex	Disease Duration Number (months)	RF titre (IU)	CCP titre (IU)	Previous DMARD (number)	Response to synthetic DMARDs	Current DMARDs	Current Biologic	% of IgG reads with <5 muts.
BP16	72	M	336	128	>340	8	No	SASP+HCQ	rituximab	32.13
BP18	54	F	141	<10	201	6	No	Mtx+HCQ	None	20.73
BP21	61	F	310	193	66	7	No	SASP+HCQ	None	20.85
BP24	70	M	109	376	>340	5	No	Mtx	Abatacept	20.16
BP31	53	F	240	10	11	4	No	Gold	None	27.89
BP33	69	F	119	359	>340	3	No	Mtx+HCQ+SASP	None	20.15
BP36	55	F	20	DNA	>340	3	No	Mtx+HCQ	None	31.95
BP40	67	F	134	890	331	3	No	LEF+SASP	None	37.96

BP41	38	F	216	417	DNA	3	No	MTX	None	40.54
BP44	64	M	372	970	DNA	5	No	LEF	None	26.67
BP9	57	M	100	10.6	75	3	No	None	Certolizumab	18.86
BP10	68	F	84	DNA	81	5	No	none	Etanercept	23.76
BP6	32	F	12	88.7	179	3	No	None	Etanercept	18.99
BP5	22	F	48	5.2	18	3	No	MTX	rituximab	62.46
BP3	39	F	84	131	142	5	No	None	Tocilizumab	96.65

This further experimental work provides further validation of the convenient, simple and reliable test for patient responsiveness to conventional DMARDs disclosed herein.

5 References

1. Visser K, *et al.* (2009) Multinational evidence-based recommendations for the use of methotrexate in rheumatic disorders with a focus on rheumatoid arthritis: integrating systematic literature research and expert opinion of a broad international panel of rheumatologists in the 3E Initiative. *Annals of the rheumatic diseases* 68(7):1086-1093.
2. Morgan C, *et al.* (2003) Contribution of patient related differences to multidrug resistance in rheumatoid arthritis. *Annals of the rheumatic diseases* 62(1):15-19.
3. O'Dell JR (2004) Therapeutic strategies for rheumatoid arthritis. *The New England journal of medicine* 350(25):2591-2602.
4. Romao VC, Canhao H, & Fonseca JE (2013) Old drugs, old problems: where do we stand in prediction of rheumatoid arthritis responsiveness to methotrexate and other synthetic DMARDs? *BMC medicine* 11:17.
5. Dorner T, Radbruch A, & Burmester GR (2009) B-cell-directed therapies for autoimmune disease. *Nature reviews. Rheumatology* 5(8):433-441.
6. Keystone E, *et al.* (2009) Rituximab inhibits structural joint damage in patients with rheumatoid arthritis with an inadequate response to tumour necrosis factor inhibitor therapies. *Annals of the rheumatic diseases* 68(2):216-221.
7. Nielen MM, *et al.* (2004) Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. *Arthritis and rheumatism* 50(2):380-386.
8. Hartung HP & Kieseier BC (2010) Atacicept: targeting B cells in multiple sclerosis. *Therapeutic advances in neurological disorders* 3(4):205-216.
9. Genovese MC, Kinnman N, de La Bourdonnaye G, Pena Rossi C, & Tak PP (2011) Atacicept in patients with rheumatoid arthritis and an inadequate response to tumor necrosis factor antagonist therapy: results of a phase II, randomized, placebo-controlled, dose-finding trial. *Arthritis and rheumatism* 63(7):1793-1803.

10. Bao J, Li T, Zhao J, & Zhang CY (2012) Cytotoxic T lymphocyte-associated antigen 4 Ig-induced asthma in the treatment of rheumatoid arthritis. *The Journal of rheumatology* 39(9):1903.
11. Ballesteros-Tato A, Stone SL, & Lund FE (2014) Innate IFNgamma-producing B
5 cells. *Cell research* 24(2):135-136.
12. Bresnihan B, *et al.* (2009) Synovial tissue sublining CD68 expression is a biomarker of therapeutic response in rheumatoid arthritis clinical trials: consistency across centers. *The Journal of rheumatology* 36(8):1800-1802.
13. Takemura S, Klimiuk PA, Braun A, Goronzy JJ, & Weyand CM (2001) T cell
10 activation in rheumatoid synovium is B cell dependent. *Journal of immunology* 167(8):4710-4718.
14. McInnes IB & Schett G (2011) The pathogenesis of rheumatoid arthritis. *The New England journal of medicine* 365(23):2205-2219.
- 15 Association WM. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA : the journal of the American Medical Association*. 2013;310(20):2191-4.
- 16 Magoc T, and Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics (Oxford, England)*. 2011;27(21):2957-63.
- 17 Glanville J, Zhai W, Berka J, Telman D, Huerta G, Mehta GR, Ni I, Mei L, Sundar PD,
20 Day GM, *et al.* Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;106(48):20216-21.
- 18 Komsta L, and Novomestky F. CRAN - Package moments. *R Package Version 012*. 2011.
- 25 19 Papoulis A, and Pillai SU. *Probability, Random Variables, and Stochastic Processes*. Tata McGraw-Hill Education; 2002.
- 20 US20120183969/ WO2012/097374
- 21 Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26(19):2460-1.
- 30 22 Smith TF, and Waterman MS. Identification of common molecular subsequences. *Journal of molecular biology*. 1981;147(1):195-7.

Claims

1. A method of determining the likelihood of a subject responding to conventional DMARDs, the method comprising:
 - a) determining the B cell hypermutation rate of said subject; and
 - b) determining therefrom the likelihood of the subject to respond to conventional DMARDs.
2. The method of claim 1 wherein hypermutation rate is determined by:
 - sequencing at least one region of interest of nucleic acids encoding the variable region of immunoglobulins (antibodies/B cell receptors) obtained from B cells from said subject to obtain a set of sequence reads;
 - comparing the determined sequence reads with corresponding germline allele sequences; and
 - determining the number of nucleotide differences present in the sequence reads and thereby determining the hypermutation rate.
3. The method of claim 1 or 2 wherein the hypermutation rate is determined for IgG.
4. The method of any preceding claim comprising providing a blood sample from said subject, preferably obtaining peripheral blood mononuclear cells from said blood sample, and more preferably obtaining B cells from said peripheral blood mononuclear cells.
5. The method of any preceding claim comprising isolating mRNA or total RNA from B cells.
6. The method of any preceding claim comprising amplifying at least a region of interest of the nucleic acids encoding immunoglobulin variable regions by PCR using suitable primers flanking the region of interest.
7. The method of claim 6 wherein the primers for use in amplifying the immunoglobulin variable regions are directed to sequences within framework region 1 and the IgG and/or IgM constant regions.
8. The method of any preceding claim wherein the complete variable region of the immunoglobulin heavy chain and/or immunoglobulin light chain is analysed.

9. The method of any preceding claim wherein libraries of sequences comprising a region of interest of nucleic acids encoding the variable region of immunoglobulins are sequenced.
- 5 10. The method of claim 9 comprising comparing sequences of a region of interest with the corresponding germline allele sequence to determine the number of nucleotide differences present.
- 10 11. The method of claim 10 wherein the number of nucleotide differences present is determined across substantially all of the nucleic acids encoding the entire variable region of immunoglobulins from B cells, typically substantially the entire variable region of both the heavy and light chains.
- 15 12. The method of claim 10 or 11 in which, for each sequence read obtained, the number of nucleotide differences in the region of interest between the sequence read and the corresponding germline V-region allele is determined.
- 20 13. The method of any preceding claim wherein the hypermutation rate is defined as a number of nucleotide differences observed per given sequence length.
14. The method of any one of claims 1 to 12 wherein the hypermutation rate is defined as the mean number of hypermutations per sequence read.
- 25 15. The method of any one of claims 1 to 12 wherein the likelihood of a patient being a conventional DMARD non-responder may be determined with reference to a threshold defined as the proportion of sequence reads which demonstrate a given occurrence level of nucleotide differences, typically an atypically low occurrence of nucleotide differences.
- 30 16. The method of any preceding claim wherein, in step b) of the method, the determination comprises comparing the determined B cell hypermutation rate with a reference value.
- 35 17. The method of claim 16, where the determined B cell hypermutation rate is lower than the reference value, the subject can be identified as a likely non-responder to conventional DMARDs.

18. The method of claim 17 wherein the reference value is a value determined from healthy controls or from RA patients who are responsive to conventional DMARDs.
19. The method of any one of claims 16 to 18 wherein, when in step b) the hypermutation rate is determined across the entire variable region, a non-responder is indicated by a mean number of hypermutations per sequence read of 17 or fewer.
20. The method of any preceding claim which comprises generating a graphical representation of the B cell hypermutation rate.
21. The method of any preceding claim which comprises the step of selecting, based upon step b), a subject for immediate or early treatment with biologic DMARDs.
22. The method of any preceding claim which is applied to subject diagnosed or suspected of having an inflammatory arthritis, especially rheumatoid arthritis.
23. A method of treating a subject having an inflammatory arthritis (suitably rheumatoid arthritis), the method comprising the steps of:
- a) providing a biological sample from a subject comprising B cells;
 - b) determining the B cell hypermutation rate of said subject;
 - c) determining therefrom the likelihood of a patient to respond to conventional DMARDs; and
 - d) administering an appropriate treatment to the subject based at least partially upon the determination in step c).
24. The method of claim 23 wherein, where the subject has a B cell hypermutation rate below a predetermined reference value the subject may be treated immediately or early with a biological DMARD.
25. An assay system for determining the likelihood of a subject to respond to non-biologic DMARDs, the assay system being adapted to carry out the method as defined in any one of claims 1 to 22.
26. The assay system of claim 25 comprising:
- one or more reaction vessels to receive a biological sample comprising B cells of said subject;

- nucleic acid isolation means to isolate nucleic acids which encode the variable region of immunoglobulins (antibodies/B cell receptors) from said B cells of said subject;
- nucleic acid amplifying means adapted to amplify at least one region of interest of the nucleic acids within the sequences encoding the variable region of immunoglobulins;
- sequencing means to determine the sequences of the amplified nucleic acids including the at least one region of interest;
- computing means configured to run suitable software to determine germline allele sequences and determine the number of nucleotide changes in the sequences obtained for the region of interest compared to the corresponding germline allele sequences.

27. The assay system of claim 26 wherein the nucleic acid amplifying means comprises suitable primers to amplify the nucleic acids including the region of interest and suitable reagents to perform PCR.

28. A kit of parts, the kit of parts being adapted for use in a method as set out above, the kit comprising suitable primers to amplify nucleic acids including the region of interest, and optionally suitable reagents to perform PCR and/or suitable probes for performing quantitative PCR.

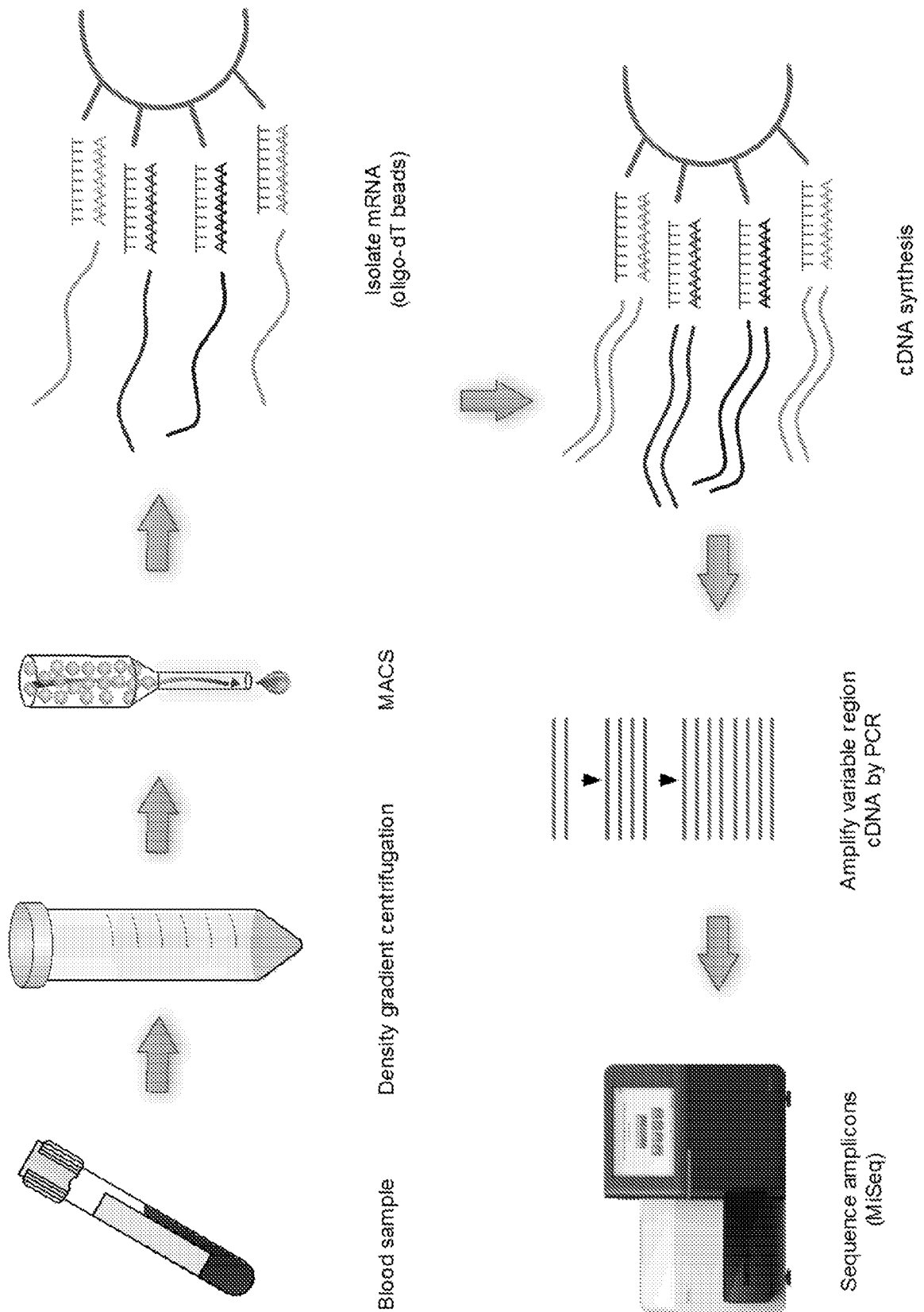


Fig. 1

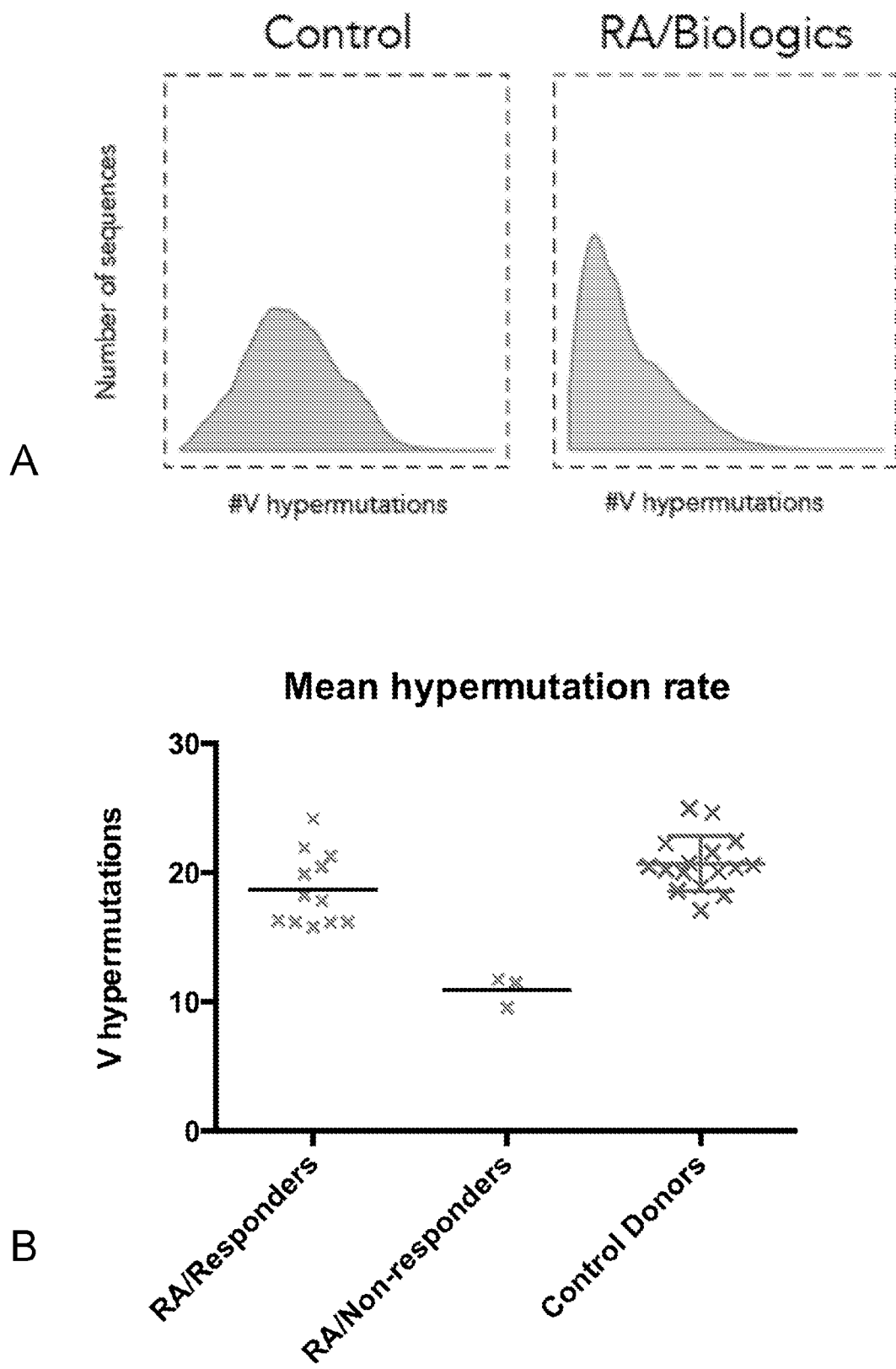


Fig. 2

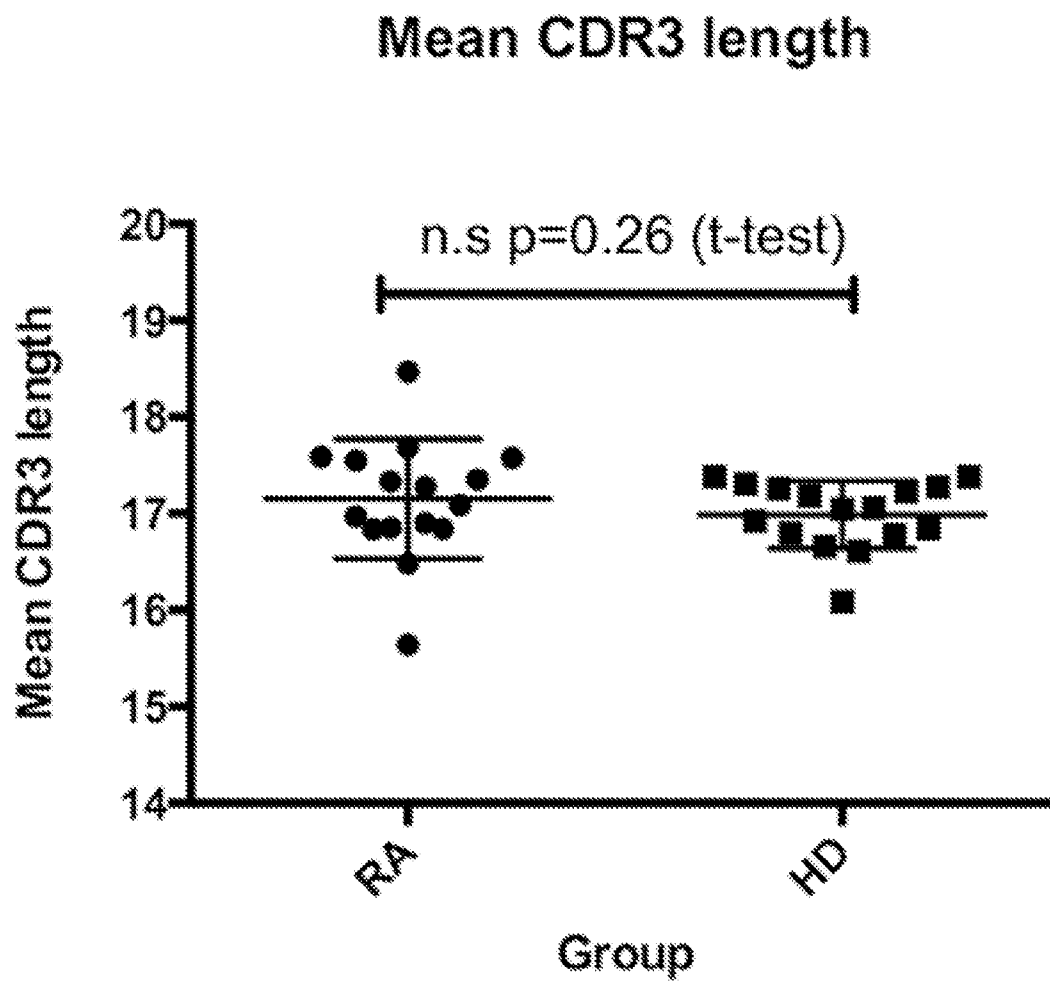


Fig. 3

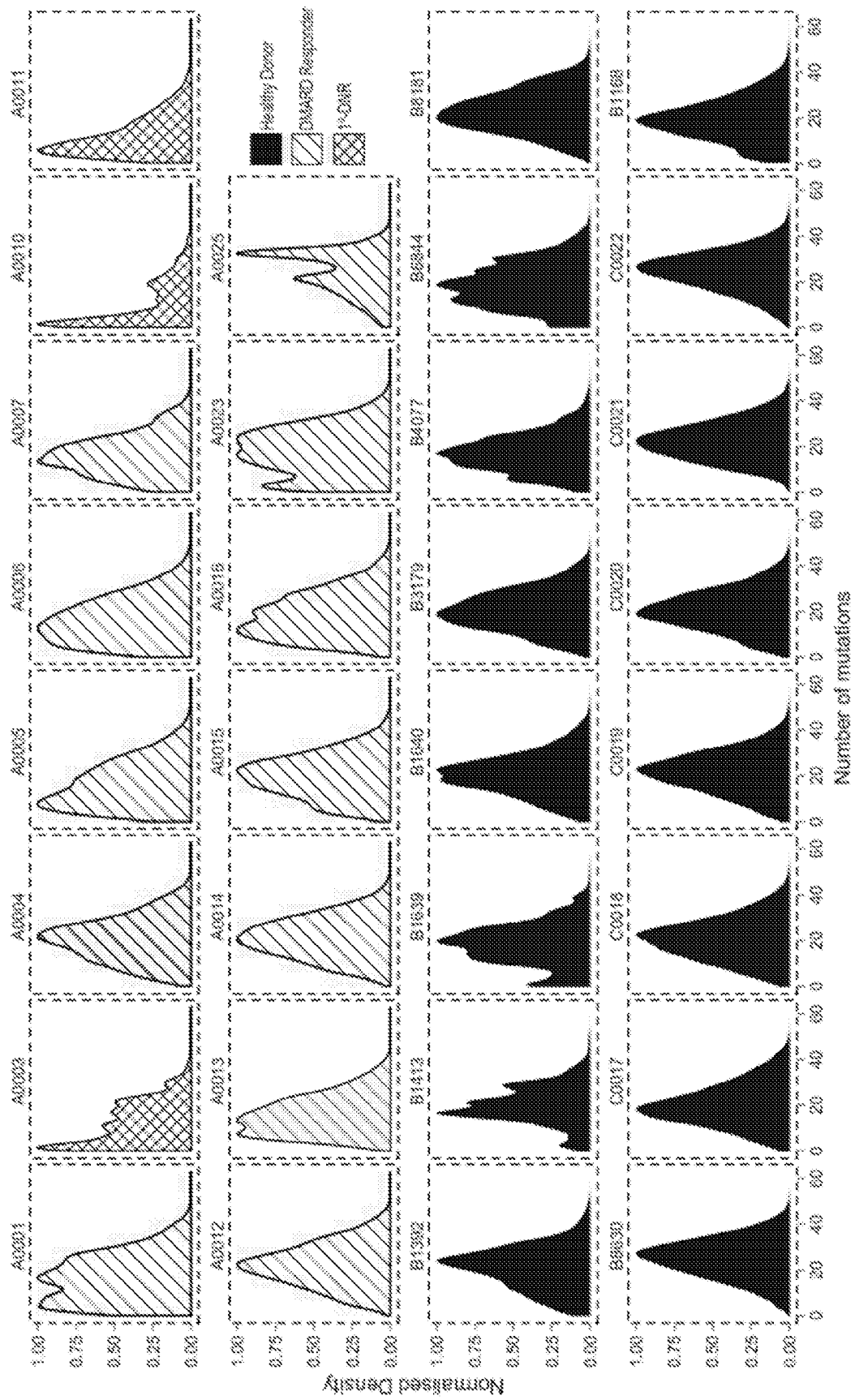


Fig. 4A

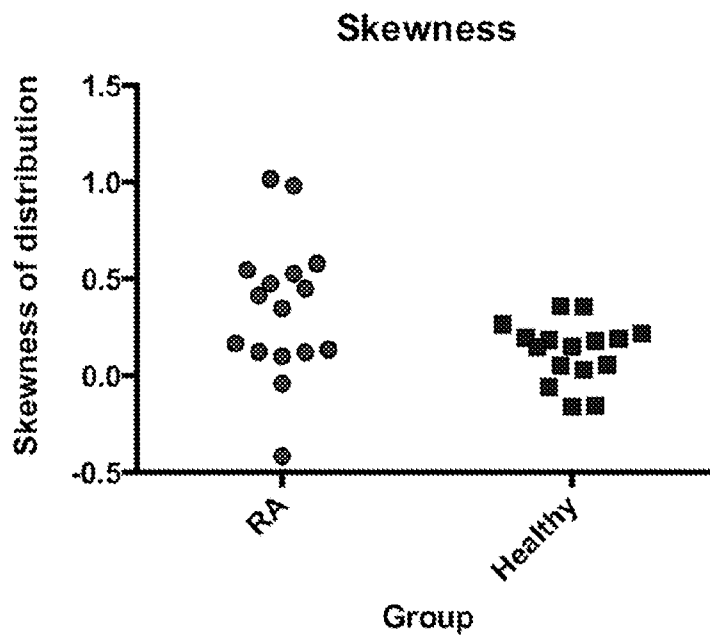


Fig. 4B

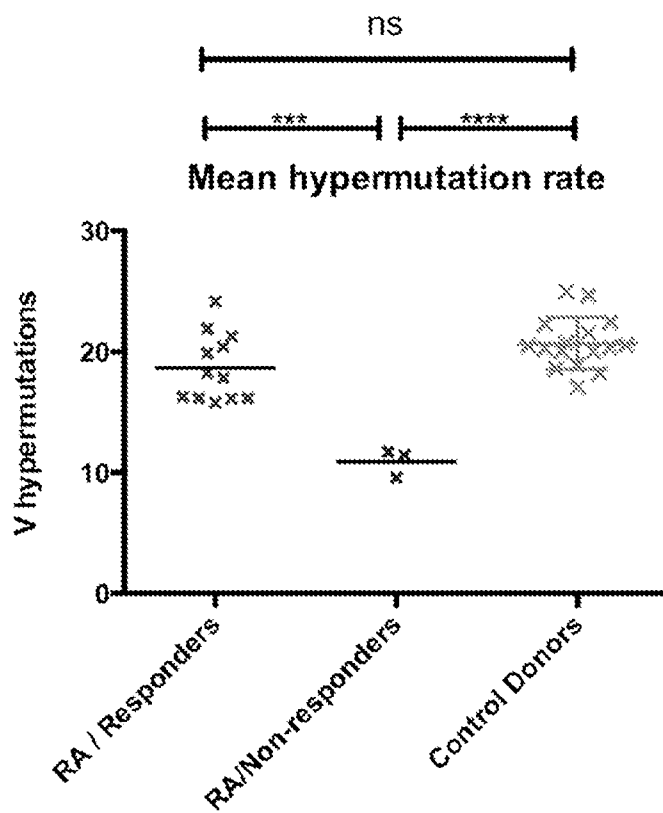


Fig. 4C

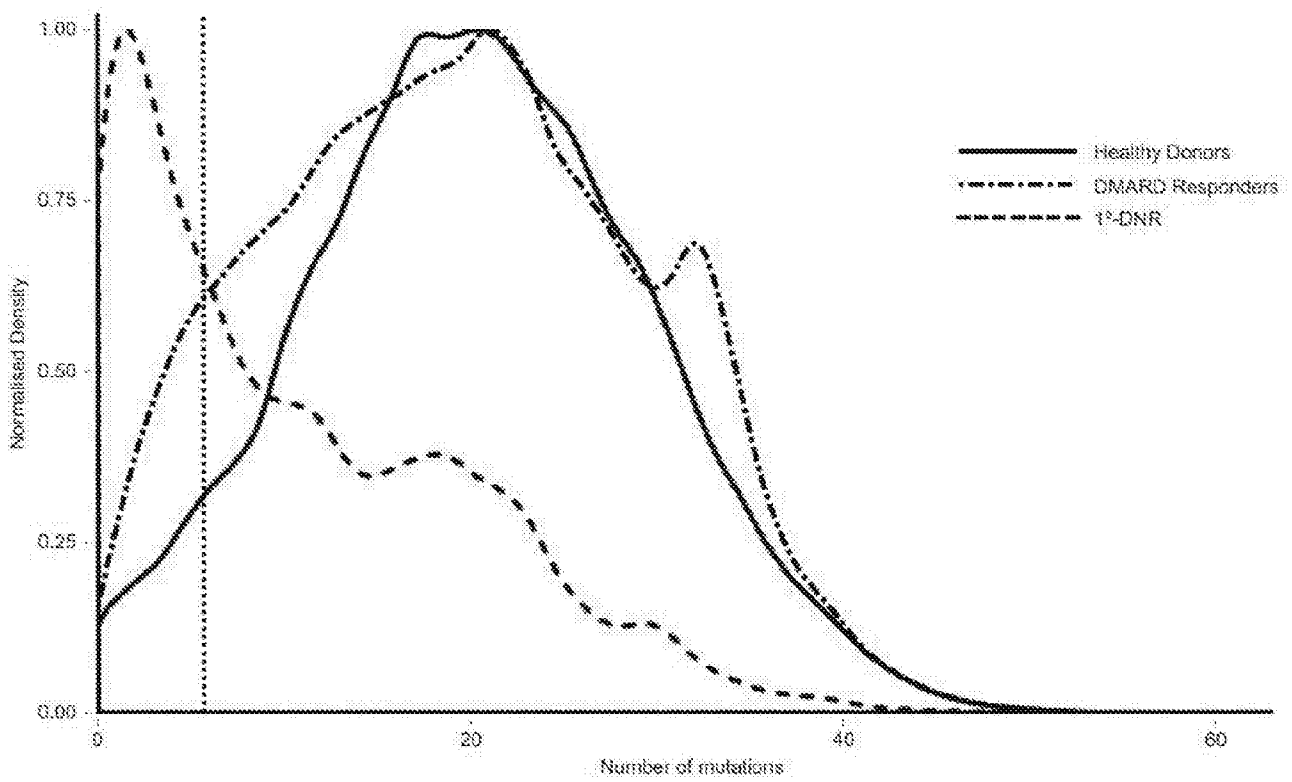


Fig. 5A

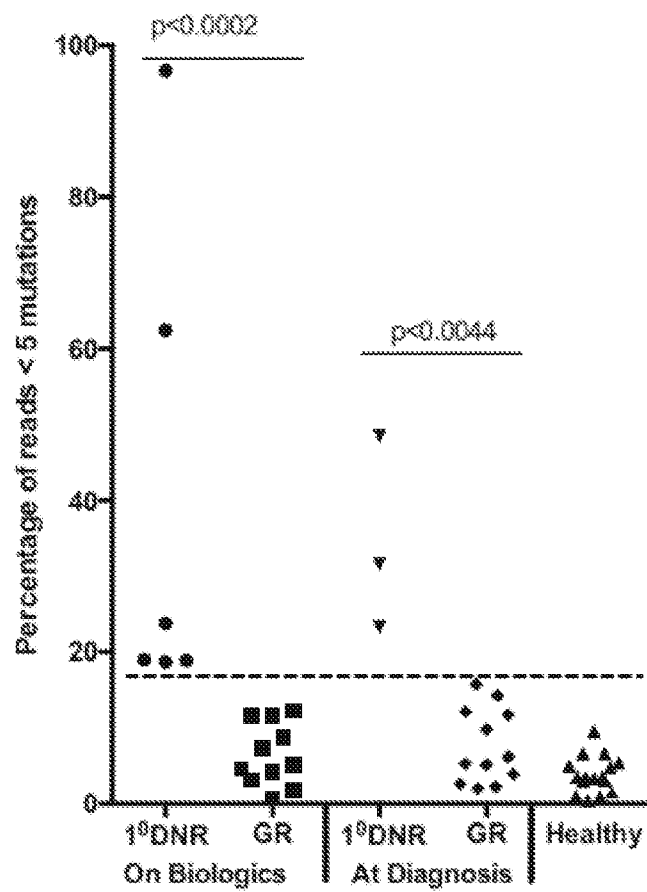


Fig. 5B

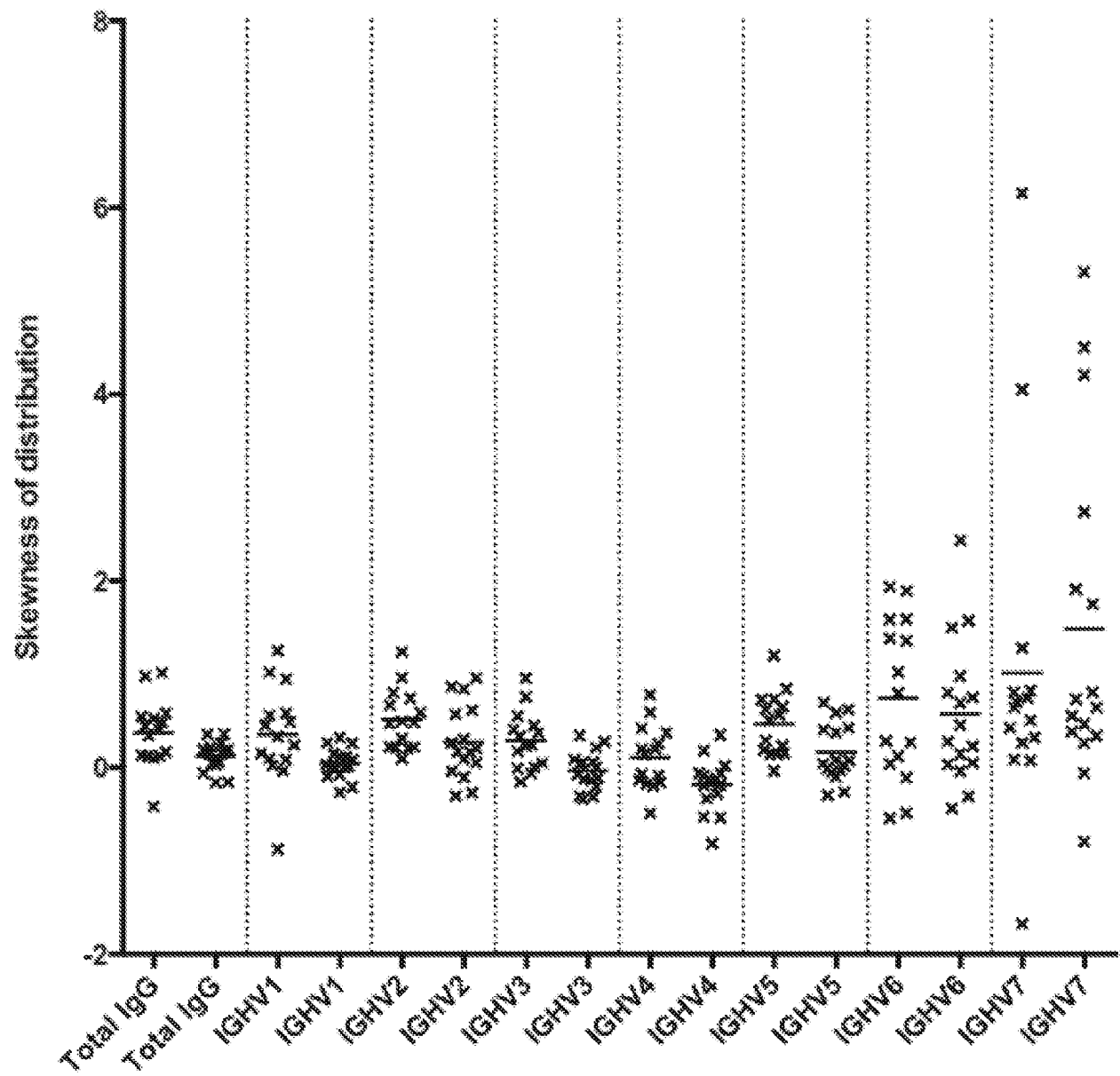


Fig. 6

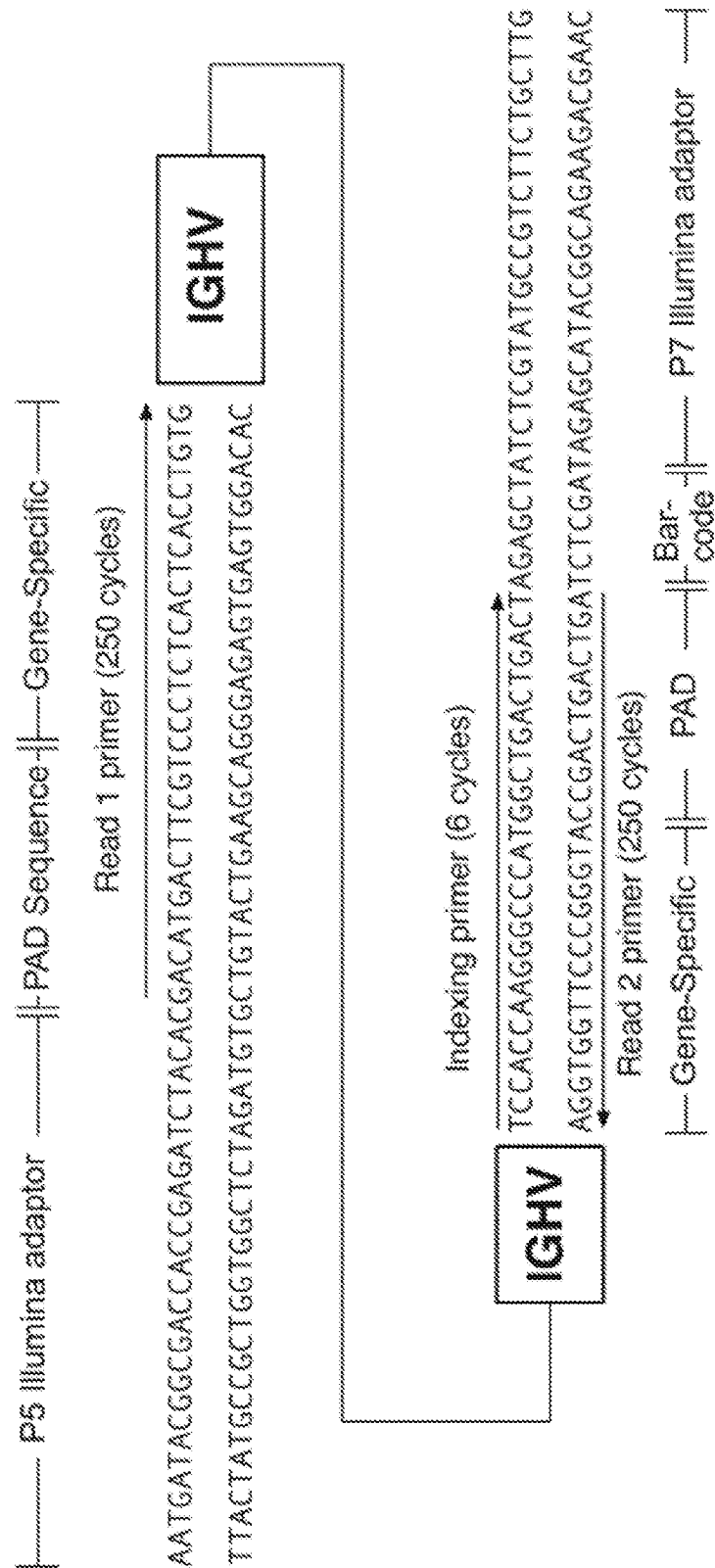


Fig. 7a

IGHV FR1 forward amplification primers	
IGHV2.0	AATGATACGGCGACCAACCGAGATCTACACGACATGACTTCGTCCTCTCACACACCTGIG
IGHV2.1	AATGATACGGCGACCAACCGAGATCTACACGACATGACTTCGTCCTKAGACTCTCTCTGTG
IGHV2.2	AATGATACGGCGACCAACCGAGATCTACACGACATGACTTCGTCCTGARACTCTCTCTGTG
IGHV3.0	AATGATACGGCGACCAACCGAGATCTACACGACACTACAGCGTAGACCCCTCACRCTGACCT
IGHV0.0	AATGATACGGCGACCAACCGAGATCTACACGACATAGACCGTTGAGTGAAGTCTCTCTGC
IGHV1.0	AATGATACGGCGACCAACCGAGATCTACACCGCGAAGCAGTCTGAGGTGAAGAAGCCT
IGHV1.1	AATGATACGGCGACCAACCGAGATCTACACCGCGAAGCAGTCTGAGGTGAAGCCT
IGHV1.2	AATGATACGGCGACCAACCGAGATCTACACCGCGAAGCAGTCTGAGGTGAAGCCT
IGHV4.0	AATGATACGGCGACCAACCGAGATCTACACCGCGTCTGTCCTGAGACTCTCTCTG
IGHV4.1	AATGATACGGCGACCAACCGAGATCTACACCGCGTCTGTCCTGAGACTCTCTCTG
IGHV5.0	AATGATACGGCGACCAACCGAGATCTACACCGTCTGAGTCTCTCTGCAAGGTTCTCTGG
IgHC reverse primer	
IGHVrev(INDEX1)	CAAGCAGAAGACGGCATACGAGATAGCTCTAGTCAGTCAGCCATGGGCCCTTGGTGG*A
IgMHC reverse primer	
IGHVrev(INDEX1)	CAAGCAGAAGACGGCATACGAGATAGCTCTAGTCAGTCAGCCGGGGGATGCACTCC*C

Fig. 7b

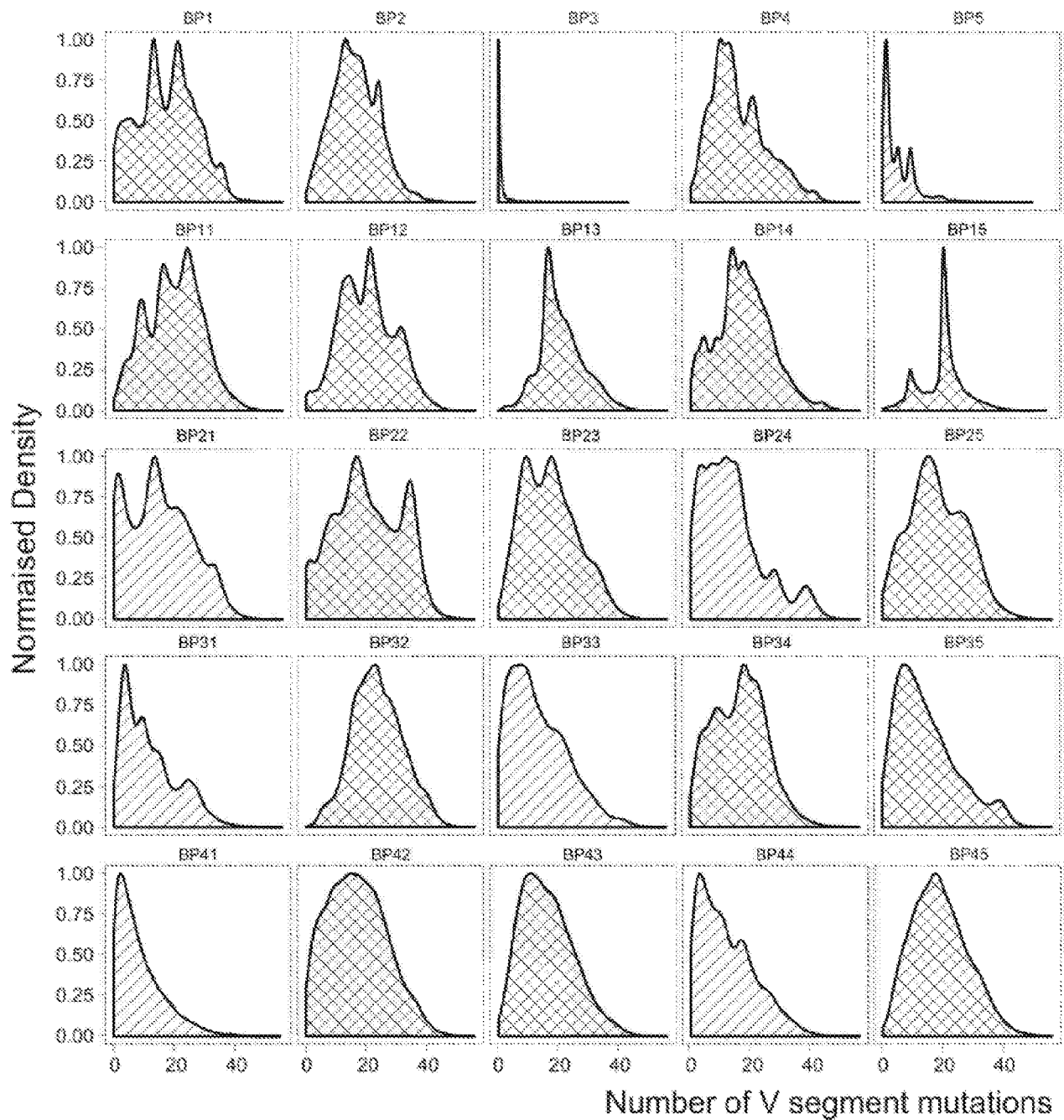


Fig. 8A

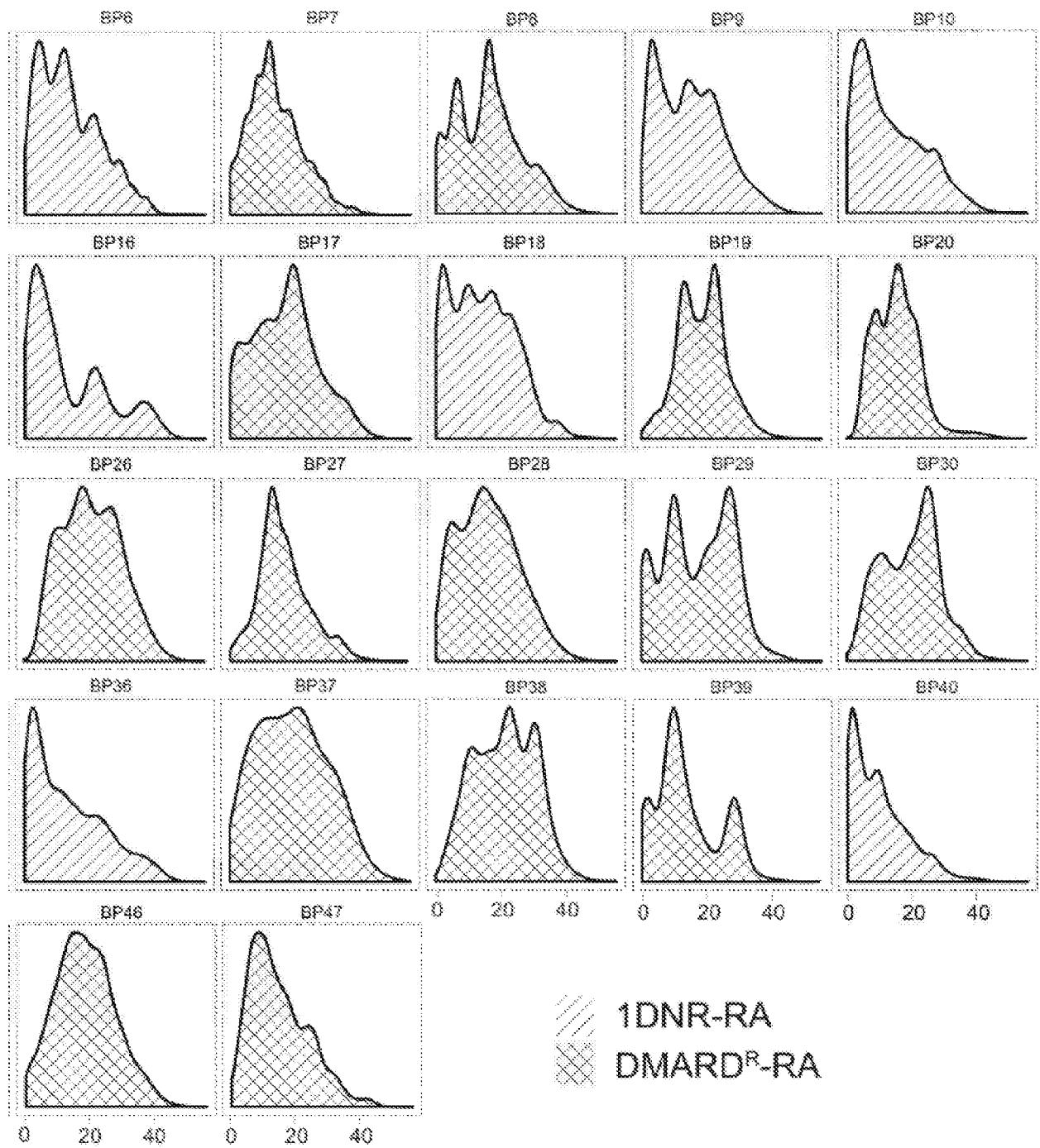


Fig. 8A continued

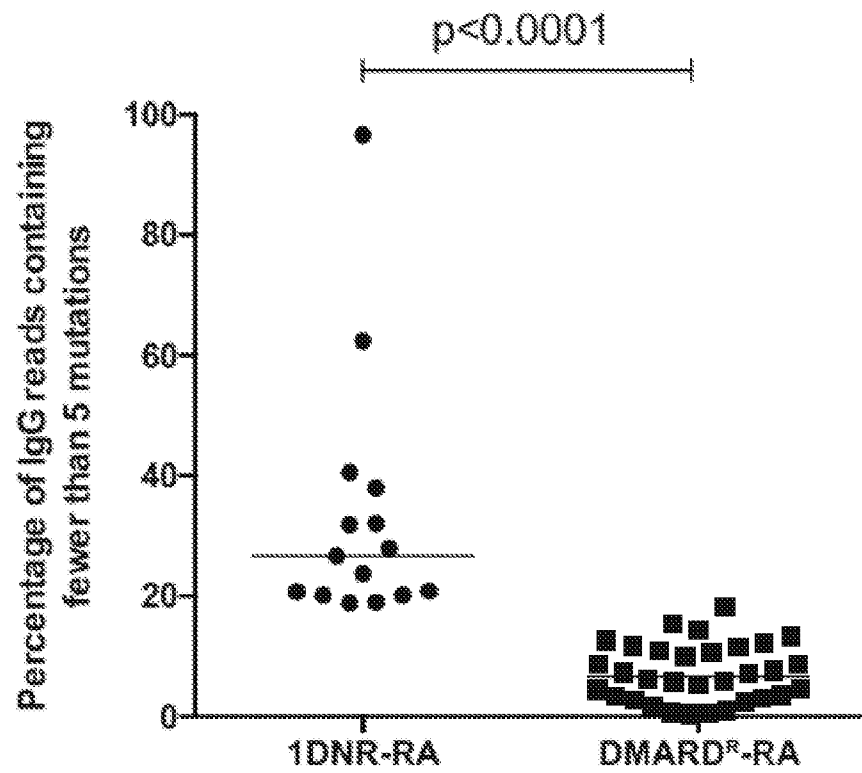


Fig. 8B

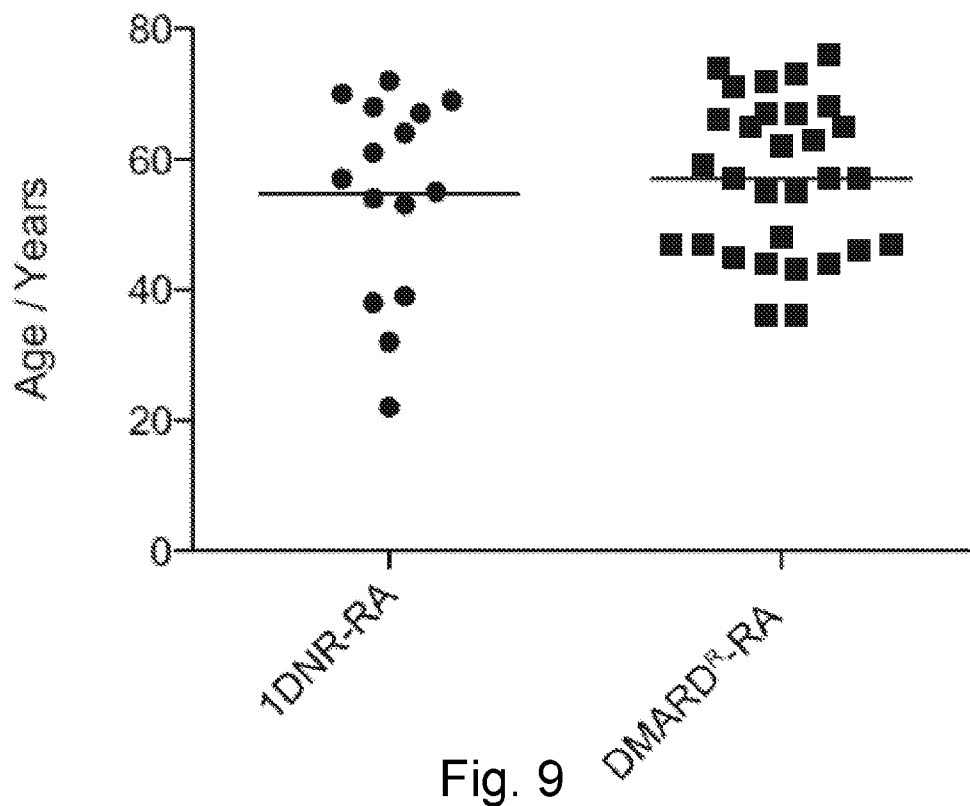


Fig. 9

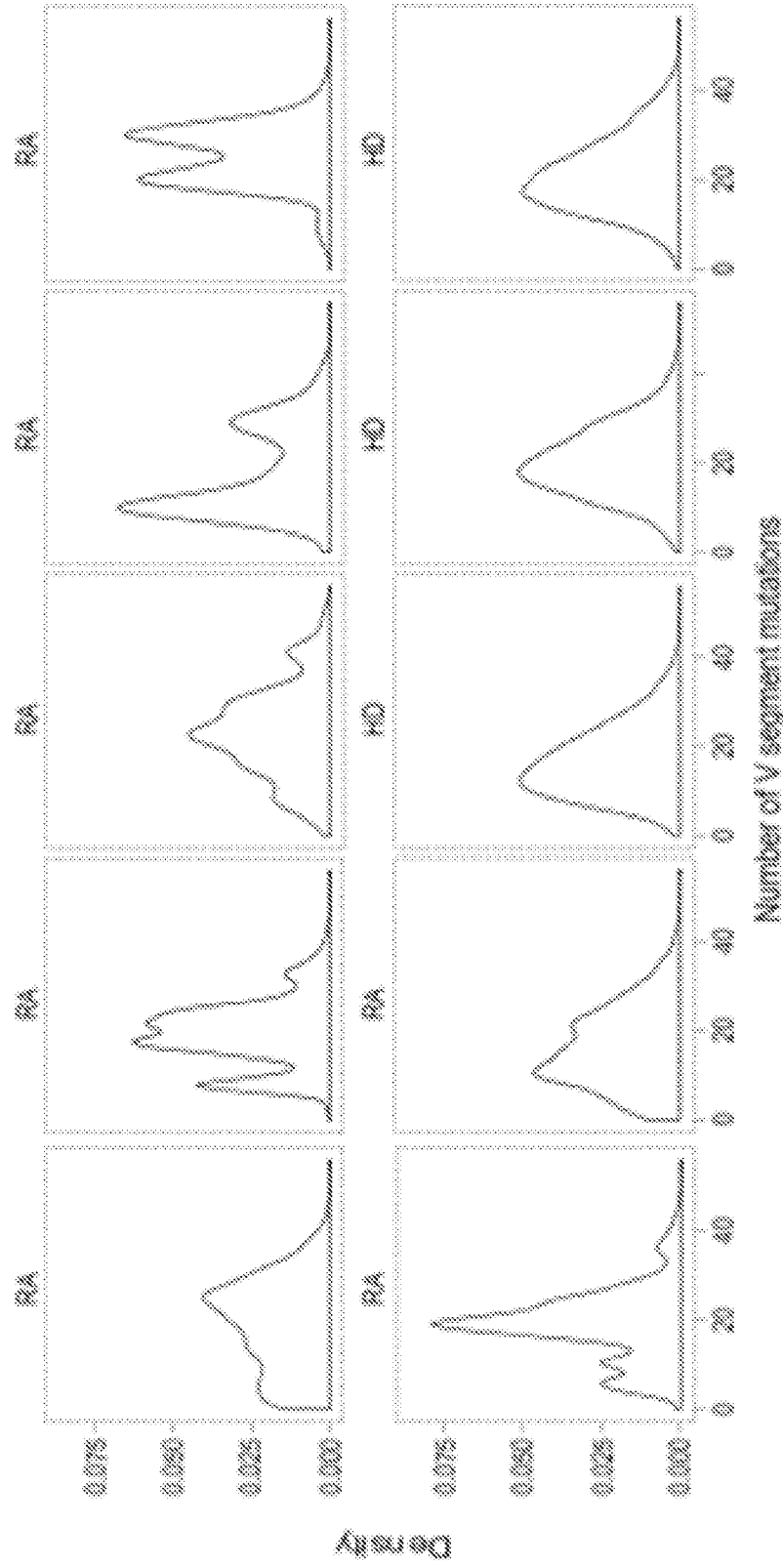


Fig. 10

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2016/051097

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2013/055595 A1 (SEQUENTA INC [US]; FAHAM MALEK [US]) 18 April 2013 (2013-04-18) the whole document	1-28
X	WO 2005/022118 A2 (PROMETHEUS LAB INC [US]; DERVIEUX THIERRY [US]; WALSH MICHAEL [US]) 10 March 2005 (2005-03-10) the whole document ----- -/--	1-28



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

13 June 2016

Date of mailing of the international search report

20/07/2016

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Mueller, Frank

INTERNATIONAL SEARCH REPORT

International application No

PCT/GB2016/051097

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ZUCKERMAN N S ET AL: "Somatic hypermutation and antigen-driven selection of B cells are altered in autoimmune diseases", JOURNAL OF AUTOIMMUNITY, LONDON, GB, vol. 35, no. 4, 1 December 2010 (2010-12-01), pages 325-335, XP027433011, ISSN: 0896-8411, DOI: 10.1016/J.JAUT.2010.07.004 [retrieved on 2010-08-19] the whole document -----	1-28

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2016/051097

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2013055595 A1	18-04-2013	US 2014256592 A1	11-09-2014
		WO 2013055595 A1	18-04-2013

WO 2005022118 A2	10-03-2005	CA 2539953 A1	10-03-2005
		EP 1668337 A2	14-06-2006
		EP 2354255 A1	10-08-2011
		US 2005112627 A1	26-05-2005
		US 2009197266 A1	06-08-2009
		US 2009197267 A1	06-08-2009
		US 2009197268 A1	06-08-2009
		US 2013131072 A1	23-05-2013
		WO 2005022118 A2	10-03-2005
