EDITORIAL

Measurable residual disease testing in chronic lymphocytic leukaemia: hype, hope neither or both?

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Leukemia (2021) 35:3364–3370; <https://doi.org/10.1038/s41375-021-01419-7>

There is a quality to quantity.

General Secretary Nikita Khrushchev

There is considerable interest in accurately quantifying remaining leukaemia cells in someone with chronic lymphocytic leukaemia (CLL), a process referred to as measurable residual disease (MRD)-testing. Reasons for wanting these data include evaluating therapy efficacy, determining whether a therapy change is needed and predicting the risk of leukaemia recurrence and other outcomes. A perfect MRD-test would accurately quantify numbers of residual leukaemia cells biologically able to cause leukaemia relapse in someone with CLL within a defined interval whilst being indifferent to leukaemia cells without this feature. Several MRD techniques focus on abnormal immune phenotype assayed by multi-parameter flow cytometry (MPFC), leukaemia-specific immunoglobulin heavy-chain gene rearrangements assayed by quantitative real-time polymerase chain reaction (qRT-PCR) with allele-specific oligonucleotides (ASOs) and mutations detected by next generation sequencing (NGS). Cells tested for MRD are typically from blood and/or bone marrow and more recently, cell-free DNA from blood. This sampling is challenging in CLL given the non-uniform distribution of leukaemia cells in bone marrow, lymph nodes, spleen and elsewhere. Most CLL studies report MRD-test results as detectable or undetectable (uMRD) rather than -positive or negative as in other leukaemias. No CLL MRD-test has perfect sensitivity, specificity, reproducibility, replicability or positive or negative predictive values (PPV, NPV). Data from clinical trials report a negative MRD-test at diverse time points during therapy identifies persons with better progression-free survival (PFS) compared with persons with a positive MRD-test after adjusting for other prognostic and predictive co-variables. However, correlations with the intended endpoint for a MRD-test, cumulative incidence of relapse (CIR), are rarely reported. The correlation between results of MRD-testing and PFS have generated interest in using MRD-test data to direct therapy and for regulatory approval. Convincing proof any intervention improves outcomes of persons with a positive MRD-test or correlate with survival are lacking and require testing in randomized trials. Refinements in and standardization/harmonization of MRD-assay platforms and test results reporting are needed to determine whether these data should be used as a surrogate endpoint in CLL therapy trials. Regardless, a MRD-test endpoint is accepted for drug approval by the European Medicines Agency (EMA) but not

yet by US Food and Drug Administration (US FDA). Although MRD-testing in CLL has advanced substantially, much remains to be done.¹

Complete response variously defined but mostly by clinical and laboratory co-variables is the goal of treating chronic lymphocytic leukaemia (CLL). This is based on the observation that amongst persons receiving therapy those achieving a complete response live longer than those with any other response type [1, 2]. However, several limitations of defining complete response by these co-variables were always obvious. One limitation is our imperfect ability to distinguish leukaemia cells from normal cells by any one or a combination of criteria. Another is variability in the distribution of leukaemia with the body, say blood, bone marrow, lymph nodes and spleen. And there are other limitations we discuss below. Given these considerations, it is not surprising many or most persons with CLL with a complete response subsequently have leukaemia recurrence or meet clinical definitions of progression. However, others do not within a defined observation interval. Why? One possibility is all leukaemia cells able to cause relapse were eliminated by therapy. Another is some or even many leukaemia cells able to cause relapse remain but simply do not do so within the observation interval. And there are other possibilities.

These limitations make it desirable to develop more sensitive techniques to quantify residual leukaemia cells, especially those able to cause relapse in someone in complete response. A question is why do we need to know this? Is it to detect some or all residual leukaemia cell(s), only leukaemia cells biologically able to cause relapse or progression within a specified interval, those which actually cause relapse within this interval or another reason? These are distinct, sometimes overlapping, but not identical goals. As such, the goal of a technique developed to detect residual leukaemia cells must be clearly defined. Several of these issues are reviewed elsewhere [3–10].

There are about $2 \times 10^6 + 12$ lymphocytes in a normal 70 kg person [11]. If we consider five percent of these lymphocytes are leukaemia cells up to $10^6 + 11$ leukaemia cells might persist in a

¹In CLL there is disagreement whether to term a situation where there are no detectable leukaemia cells and normal bone marrow function by conventional methods complete remission or complete response. There is also controversy over whether to do MRD-testing in someone with complete remission or response. Herein, we use the term complete response. We also follow the US FDA recommendation to restrict MRD-testing to persons in complete response. Consequently, the outcome measures we consider are cumulative incidence of leukaemia recurrence (CIR), event- and relapse-free survivals (EFS and RFS) and survival. We are aware of the ambiguity in the term progression-free survival in CLL because it can refer to persons with previously untreated CLL who then receive and/or require therapy (often a subjective decision) from a person with a complete or partial response who progress and may receive and/or require therapy.

person declared to have a complete response. However, this estimation of residual leukemia cells and attendant implication(s) is based on several likely incorrect assumptions: (1) every residual leukaemia cell can cause relapse; (2) leukaemia cells able to cause relapse in a defined interval are uniformly distributed in the blood, bone marrow and elsewhere; (3) there is no error rate of estimating the proportion of leukemia cells when evaluating relatively few cells; and (4) a small sample volume of blood or bone marrow is taken from someone will be representative of the quantity and distribution of leukaemia cells in that person.

METHODOLOGY CONSIDERATIONS

One definition of a perfect MRD test is it can accurately identify the smallest population(s) of leukaemia cells in someone with CLL with a complete response defined by International Workshop on Chronic Lymphocytic Leukaemia (iwCLL) criteria which, left untreated, cause leukaemia relapse within a prescribed interval whilst being indifferent towards residual leukaemia cells without this biological capacity. For the clinical performance of any MRD test, the theoretical maximal sensitivity and specificity of an assay to detect such residual leukaemia cells together with reproducibility and replicability are important as are practical considerations regarding sampling details (site, volume, timing, frequency etc.) and results interpretation.

Some leukaemias such as chronic myeloid leukaemia (CML) and acute promyelocytic leukaemia (APL) are associated with canonical mutations, *BCR::ABL1* and *PML::RARA*. Almost everyone with these leukaemias has these mutations which are *necessary and sufficient* to cause leukaemia [12, 13]. In this ideal setting quantification of transcripts of these genes using quantitative real-time polymerase chain reaction (qRT-PCR)-based MRD-testing is of considerable clinical value [14–19]. However, even in this ideal situation most persons with a negative MRD-test, especially in CML, have residual leukaemia *stem cells* [20]. Moreover, about one-half of persons with CML with a consistently negative highly sensitive MRD-test relapse quickly relapse when tyrosine kinase-inhibitor (TKI) therapy is stopped [15, 21]. However, detecting MRD in CLL is far more challenging given the absence of a canonical causative mutation and the considerable genotypic heterogeneity. Consequently, there is considerably more complexity, inaccuracy and imprecision to detecting MRD in persons with CLL compared with persons with CML and APL. Perhaps a better model for MRD-testing in CLL is acute lymphoblastic leukaemia (ALL) where despite lack of a canonical mutation(s), results of MRD-testing are useful to predict outcomes and direct therapy despite substantial false-positive and negative test results.

Another important consideration is sampling error. Typically, we sample a small volume of blood and/or bone marrow. 1st, there is the limitation of non-random distribution of leukaemia cells, especially in relatively inaccessible sites such as lymph nodes and spleen. 2nd, as we previously reported when there are relatively few non-uniformly-distribute residual leukaemia cells in someone, typically $< 10E + 4$ there is a substantial probability a small blood or bone marrow sample will not contain a leukaemia cell [22]. In this instance, no MRD-test, regardless of sensitivity can be a true-negative. Put otherwise, at low residual leukaemia cell numbers sample volume rather than MRD-test sensitivity is rate-limiting. This consideration is often ignored by persons trying to develop increasingly sensitive MRD-tests.

MULTI-PARAMETER FLOW CYTOMETRY

The diverse methods to quantify MRD in CLL rely on the phenotype or genotype of the leukaemia cells. Multi-parameter flow cytometry (MPFC)-based MRD-tests focus on the surface immune phenotype of the leukaemia cells [2, 23–26]. They operate by detecting cell population(s) consistently and reliably

different from an antigen-expression pattern of normal or regenerating cells of similar lineage and maturation stage. Targetable deviations include cross-lineage expression, over-expression, reduced or absent expression, asynchronous expression and others. The most common target phenotype of CLL is CD19/CD5 co-expression but this also requires proof of clonality by immunoglobulin light-chain restriction. The European Research Initiative in CLL (ERIC) recommends a 6 antibody panel (CD19, CD20, CD5, CD43, CD79b and CD81) which under appropriate circumstances can reliably detect $10E-5$ leukaemia cells in a sample [27]. Other potential targets for testing in MPFC may be identified in the future.

Advantages of MPFC-based MRD-detection include wide applicability, ease of quantifying abnormal cell population(s), relative sensitivity ($10E-4$ to $10E-5$ i.e. 1 in 10,000 or 100,000 cells), rapid turn-around, ability to distinguish live from dead cells and low cost. However, beyond limitations common to all MRD-testing discussed below, there are other limitations inherent to MPFC MRD-testing: (1) it is possible not all leukaemia cells have the targeted abnormal genotype or phenotype; (2) phenotypes may change over time with gains/losses of specific abnormalities or patterns of abnormalities because of disease evolution, sub-clonal selection, cell-cycle progression and/or effects of therapy; (3) sensitivity of MPFC-based MRD-testing is less than an optimized PCR-based MRD-testing (discussed below); (4) MPFC-based MRD-detection is not uniform between people with CLL because the ability to identify abnormal cells depends on the degree residual leukaemia cells differ from normal cells or from residual leukaemia cells which are biologically unable to cause relapse or progression or which fail to do so during the observation interval; (5) using MPFC appropriately requires considerable expertise and experience; analysis and data interpretation have some subjectivity and therefore operator-dependent biases making assays challenging to harmonize across laboratories; and (6) most but not all laboratories use the antibodies to identify target antigens for MRD detection by MPFC. Many of these problems are reducible with standardized laboratory procedures including sample processing and instrument settings, single tube approaches with a pre-configured and stable assay, similar antibody panel, automated interpretation software, central review and continuous quality assessment. Absent these actions one should be cautious interpreting data from a non-standardized laboratory, commercial or university-based.

MOLECULAR TECHNIQUES

Polymerase Chain Reaction (PCR)

Leukaemia cells from each person with CLL have a unique *IgH* gene locus derived from combining variable (V), diversity (D) and joining (J) gene segments (*IGHV-IGHD-IGHJ*) rearrangement which can be amplified using consensus primers for *IGHV* and *IGHJ* at the 5' and 3' of the re-arranged region. A clonal PCR product of the same size as detected at diagnosis allows identification and quantification of residual leukaemia cells. MRD evaluation by PCR is based on the identification of a person-specific molecular target on which sequence primers and probes are designed. One strategy is to use an allele-specific oligonucleotide (ASO) combined with a 3' *IGHJ* reverse consensus primer and a fluorescent probe between the primers. Although a PCR test with a person-specific molecular target approach has high sensitivity and specificity in MRD detection for CLL, it is less often used for assessing MRD because (1) need for a diagnostic sample; (2) labor intensive and time-consuming process of designing and testing subject-specific primers; (3) potential emergence of a new clone or sub-clone resulting in a false-negative MRD-test result; (4) low levels of non-specific amplification of normal or pre-leukaemia cells can result in a false-positive MRD-test result; and (5) cost [24]. Also, some persons may have more than 1 CLL clone when tested.

Next Generation Sequencing (NGS)

NGS technologies have been applied in monitoring MRD in CLL [28]. One strategy is to use multiplex PCR followed by sequencing to identify and quantify the signature immunoglobulin heavy and light-chain rearrangement of in CLL cells. Compared to the qRT-PCR with ASO, NGS-based assay uses consensus primers without the need of customized patient-specific primers resulting in a broader application in the detection of MRD in CLL.

Increased understanding of the genomic landscape of CLL has prompted considerable interest in developing MRD-tests based on detecting and quantifying somatic mutations. This task is complicated by several factors: (1) genetic clonal heterogeneity at diagnosis with evolution over time and possible emergence or selection of small sub-clone(s) at relapse; (2) error rates intrinsic to most conventional NGS techniques allow only for low sensitivity detection of mutations; (3) understanding clonality in any sample may be limited by the depth of sequencing and algorithms used for mutation calling; (4) mutated genes such as *ASXL1* can be detected in some healthy people without haematological abnormalities [29, 30]. This is especially so in older persons about the same age as most persons with CLL; and (5) some genetic abnormalities persist in persons in long-term response, possibly because of residual preleukaemia cells or expansion of normal cells with age-related somatic mutations. Despite these limitations, technical advances and increasingly sophisticated understanding of the clonal somatic mutation hierarchy in CLL means NGS-based approaches may play an increasing role in MRD-testing in the future. However, this approach needs to address the issue that not all mutations have equal biological consequences in CLL and how this computationally demanding and time-consuming technology can be brought into routine clinical practice. ERIC has published data on the use of molecular techniques to detect *TP53* mutations and comparative analyses of targeting sequencing panels to detect mutations in CLL [31, 32]. Examples of the complexity of using NGS to predict relapse hazards are interaction between *TP53* variable allele frequencies (VAFs), *IGHV* mutation and therapy type [32]. The implication is isolated analyses of mutation topography are unlikely to be accurate predictors of outcomes.

One approach to increase the accuracy and precision of MRD-testing in CLL is orthogonal testing using MPFC and parallel high-throughput sequencing resulting in good linearity to a detection limit of $10E-6$ [27]. This approach was recently tested in several CLL clinical trials [33, 34].

Recently, US FDA approved the ClonoSEC® assay which uses NGS to detect MRD in persons with CLL but has not approved results of MRD-testing as a surrogate endpoint for drug approvals (vide infra) [27]. This assay can detect leukaemia cells derived from the initial CLL clone but also emerging sub-clones of leukaemia cells making it especially attractive.

LIMITATIONS OF MRD-TESTING

As intellectually appealing as the concepts underlying relying on results of MRD-tests to predict outcomes and direct therapy in CLL, there are practical and logistical discordances between theory and practice. No current MRD-test has perfect sensitivity or specificity to accurately predict risk of relapse or relapse hazard at the cohort level or, at the more clinically relevant, individual level where therapy decisions are made. Some persons with a negative MRD-test progress during the observation interval (false-negatives) whereas others with a positive MRD-test do not (false-positives) and are cured, at least operationally cured meaning the leukaemia does not recur during their remaining lifetime or during the observation interval. Why? Besides insufficient sensitivity of the assay [28], one reason for false-negative MRD-test results is an inconsistent expression of the designated target(s) or marker(s) of the MRD-test on leukaemia cells. This is true of MRD-tests using

MPFC and molecular techniques. Although CLL is typically clonal phenotypically and genotypically, spontaneous and therapy-induced clonal evolution and selection mean a target of a previously informative MRD-test may be less useful at another time point even in the same person. Other reasons include inhomogeneous distribution of leukaemia cells in the body, inadequate sampling/sample volume, interactions of leukaemia cells with the micro-environment and the host immune response we discuss above and elsewhere [35].

Another important issue is the failure of MRD-testing to capture the biological potential of residual leukaemia cells to cause relapse. Relapse hazard correlates not only with the numbers of residual leukaemia cells but with their growth kinetics, especially when the observation interval is brief. The latter is influenced by the biology of leukaemia such as *IGHV* somatic hyper-mutation, *TP53* mutation, and potential differences in the host's micro-environment and immune response [2, 36]. Moreover, MRD-test results are an imperfect predictor of relapse hazard even when known predictive co-variables are adjusted for likely reflecting the impact of *latent* co-variables (co-variables currently unknown but potentially knowable) and stochastic events. An example of this potential discordance is the result of Bruton tyrosine kinase-inhibitor (BTKi) therapy which prolongs PFS but rarely achieves an MRD-negative state. This contrasts, for example, with therapies that include venetoclax where achieving a MRD-negative test is more common. The implication is the predictive accuracy of a MRD-test result can be different in different therapy settings and with different outcomes endpoints. Consequently, the *gold standard* in comparing therapies in person in complete response must be the CIR, EFS, RFS or survival determined in a randomized controlled trial. This highlights the potential danger of defining complete response based only on MRD-test result or approving new drugs based on MRD-test data only. We discuss the issue of surrogacy for MRD-tests below.

The frequency of MRD-testing may also impact test performance [37]. False-positive and -negative MRD-test results are more likely with single time-point measurements [38]; re-testing can increase MRD-test prediction accuracy. Sometimes repeat MRD-tests are discordant with no interval intervention which can be bi-directional: negative to positive or the converse. Discordances have many explanations but sampling variability (having at least one leukaemia cell in the sample or not) is important especially when there are few non-uniformly-distributed leukaemia cells. Requiring concordant results to declare someone MRD-positive or -negative increases specificity but decreases sensitivity [22]. Orthogonal validation is useful for discordant instances but does not resolve sampling error unless sampling is repeated. Sequential monitoring is a useful strategy to increase sensitivity if changes in MRD-levels such as increasing proportions of leukaemia cells identified by MPFC are used as the read-out. The optimal interval and duration of sequential MRD-testing is unknown and may depend on co-variables such as time from achieving a complete response, leukaemia cells biology and others. Given the potential risk of harm from unneeded additional therapy prompted by a false-positive MRD-test we suggest a confirmatory MRD-test result within a reasonable interval, perhaps 2–4 weeks, from the first positive MRD-test in someone previously MRD-test negative if an intervention is planned [39].

A MRD-test result can be false-positive because of the assay (e.g. technical errors or laboratory contamination) or failure to reflect eventual relapse- or progression because of eradication of biologically-important leukaemia cells with subsequent therapy (ies) or, possibly by immune-mediated anti-leukaemia effects, short observation interval or death from causes other than CLL within the observation interval. Just as the detection of cytogenetic or genetic abnormalities in an otherwise healthy person is not intrinsically disease-defining, detection of some abnormalities targeted by MRD-testing other than a clonal

population of B-cells in someone who has completed therapy does not necessarily indicate residual CLL or relapse or progression risk, a principle more generally seen in solid cancers. There is also the problem of leukaemia cells in un-sampled sites such as lymph nodes and spleen. Lastly, although results of MRD-testing using NGS when done on stored samples often correlate with clinical outcomes there is considerable heterogeneity [28, 33, 40]. Sample for MPFC-based MRD-assays should be done within 48 h. This limitation may be mitigated in part by standardization, formal proficiency testing and automated analyses [41]. Cross-study comparisons of MRD-testing are limited by the lack of an independent reference standard for inter-laboratory proficiency testing. This has been done in CML and needs repeating in CLL although it is difficult to compare DNA- and RNA-based MRD-testing.

MRD NOMENCLATURE

Nomenclature for describing results of MRD-testing in general, and especially in persons with CLL, is problematic. In other leukaemias many MRD-test results are reported as MRD-negative or MRD-positive. Knowing how to interpret results reported in this manner is impossible without details of sampling, specificity, sensitivity, reproducibility and replicability. The same limitations to interpretation apply to what appears to be the preferred term for reporting CLL MRD-test results: undetectable MRD (uMRD). We recommend reporting MRD-test results in CLL with data on sensitivity, specificity, PPV, NPV and the need for confirmatory repeat testing within an appropriate interval. Clinicians should be informed of uncertainty in reporting results when reduced to a binary and other caveats we discuss above, especially the lack of convincing data an intervention based on a positive MRD-test alters a person prognosis.

MRD-TESTING TO PREDICT RELAPSE OR PROGRESSION

Many recent studies in persons with CLL complete response report correlations between MRD-test results and clinical outcomes such as PFS and survival in persons receiving chemo-immuno-therapy or a haematopoietic cell transplant [1, 42, 43]. However, this seems not so in persons receiving BTK-inhibitors where there are low rates of complete responses and MRD-test-negativity but good PFS compared with other therapies with higher rates of complete response and MRD-negativity indicate limitations of using MRD-test data for drug approvals (vide infra).

Comparisons between studies with MRD-test data as the endpoint are complicated by differences in accuracy (sensitivity and specificity), timing, frequency of MRD-testing, precision (reproducibility and replicability) and other co-variables. Unsurprisingly, the probability of having a positive MRD-test in conventionally defined complete response is associated with the cytogenetic/molecular prognostic risk category and with other adverse clinical prognostic factors such as older age and receiving/failing prior therapy(ies). Nevertheless, multi-variable Cox regression analyses consistently indicate MRD-test results independently correlate with clinical outcomes for some but not all therapies. Consequently, therapy-response measured by MRD-testing may be a stronger predictor of CIR and/or PFS compared with pre-therapy co-variables but this has not been critically tested. At the cohort level, MRD-test results refine risk-stratification beyond that provided by conventional prognostic and predictive co-variables and are being included in future time-dependent prognostic and predictive models [36, 44].

Although the predictive value of MRD-test results at one-time point in predicting PFS is convincing, data on the relevance of results of MRD-test result kinetics are mixed and additional studies are needed to clarify how data from sequential MRD-testing are best used to predict outcomes and therapy decision-making.

Although current MRD-tests are an important tool for estimating relapse hazard in persons with CLL in a conventionally defined complete response some data caution against over-emphasizing results of MRD-tests to predict subsequent risk and timing of relapse or progression at the individual level especially when MRD-testing is done only once. Although many studies report results of MRD-testing to correlate better than complete response with PFS the question is to what extent current criteria of complete response will be used in the future. We believe that before this happens there should be more convincing data a MRD-test-based definition of complete response correlates with outcomes, especially CIR and RFS and, if possible, survival, in diverse CLL populations, disease states and therapy settings.

MRD-TESTING IN PERSONS WITHOUT A COMPLETE RESPONSE

The MRD-test was developed to determine whether there were remaining leukaemia cells in someone with acute myeloid leukaemia (AML) in histologically defined complete remission and, if yes, how many [45, 46]. This objective ignored whether these residual leukaemia cells had the biological ability to cause leukaemia recurrence during a defined observation interval. If the goal of MRD-testing is to quantify numbers of residual leukaemia cells the biological endpoint should be CIR. Other endpoints such as EFS, RFS, PFS and survival are confounded by events unrelated to numbers of biologically relevant residual leukaemia cells such as infection, arterio-sclerotic cardio-vascular disease, accidents etc.

We acknowledge that there are many reports of correlations between results of MRD-testing and endpoints other than CIR. However, as discussed, this is not the conceptual objective of an MRD-test. Although CIR is not a useful endpoint in CLL therapy, it is the logical biological endpoint of a clinically useful MRD-test designed to quantify numbers of residual leukaemia cells. So, why should the results of MRD-testing correlate with these other endpoints? As indicated, a small blood or bone marrow sample will only have a reproducibly positive MRD-test result when someone has many widely distributed leukaemia cells. So, someone might have an event or leukaemia progression not because of numbers of residual leukaemia cells but because of drug resistance, non-compliance, the emergence of a sub-clone and other considerations correlated or not with whether they have residual leukaemia cell. Finally, if PFS of persons with a complete response as defined by iwCLL criteria correlates with results of MRD-testing it means the MRD-test is a predictive covariate which may be unrelated to numbers of residual leukaemia cells biologically able to cause leukaemia recurrence within an observation interval unless, of course, the complete response designation is a false-positive. It is likewise conceptually difficult to understand how PFS in persons with a negative MRD-test can be similar in persons with or without a complete response as reported [47]. In this instance, the MRD-test is best considered a potentially useful predictive biomarker but not a measure of numbers of biologically relevant residual leukaemia cells. It is important to recall that in all the clinical settings we discuss we are considering correlations that do not necessarily reflect cause-and-effect.

MRD-TESTING AS A TOOL FOR DECIDING THERAPY

The close association between MRD-test results and relapse hazard has generated substantial interest in using results of MRD-testing to direct therapy decisions in persons with CLL. Inherent in a strategy of giving more, more intensive or different therapy(ies) to someone with a positive MRD-test is the hope the intervention(s) will decrease relapse hazard and improve PFS, RFS and/or survival. Several studies report most persons with CLL with a positive MRD-test result, but not necessarily because of (vide infra) large numbers of residual leukaemia cells, progress soon after the

determination [33, 47]. Inherent in the strategy of giving different therapy to someone as soon as they have a positive MRD-test is the unproven belief acting before there is loss of conventionally defined complete response will improve outcomes. For example, in the context of AML median interval from a positive MRD-test to histological relapse is 3–4 months [45]. This interval is likely to be considerably longer in CLL but this is untested. Also, considerable data in AML indicate whilst a positive MRD-test is associated with a high CIR and worse survival, there are no convincing data outcomes of persons who are MRD-test-positive randomized to receive different interventions change posttransplant outcomes. For example, conclusions of two RCTs comparing posttransplant CIRs between subjects who are MRD-test-positive receiving reduced-intensity pretransplant conditioning (RIC) or conventional transplant are contradictory [48, 49]).

The concept of MRD-test-directed therapy has precedent in persons with acute lymphoblastic leukaemia (ALL) where data from several non-randomized prospective trials indicate better outcomes with this strategy [50–53]. However, the results of randomized studies are contradictory [54–56]. In APL results of MRD-testing are only available from non-randomized studies and suggest therapy based on MRD-test results predict CIR [39, 57, 58]. There are no data from randomized controlled trials proving the efficacy of MRD-test directed therapy in AML in the context of CIR.

There are many potential biases complicating the interpretation of results of non-randomized clinical trials which claim therapy decisions based on MRD-test data result in better outcomes in CLL. Moreover, a correlation between a positive MRD-test and relapse hazard should not be assumed to be *causal*, namely, relapse may not result directly from the number of residual leukaemia cells detected in the MRD-test but because the test identifies persons most likely to relapse much like an abnormal exercise ECG correlates with risk of a subsequent cardiovascular event but does not cause it. Thus, although switching from MRD-test-positivity to -negativity is a reasonable immediate therapy goal, only randomized trials can definitively determine whether converting someone to MRD-negativity with additional therapy is associated with reduced hazard of relapse a lower CIR or longer EFS, PFS or survival.

An important variable in evaluating results of an MRD-test is physician and patient tolerance for false-negative and -positive MRD-test results. For example, if a therapy intervention prompted by a positive MRD-test is associated with few adverse events one might be tolerant of false-positives. However, there should be less tolerance for a false-positive MRD-test when the consequences of therapy are potentially adverse. Often it will be difficult to achieve equipoise between risks of relapse and adverse events.

The bottom line question to these considerations is whether outside the context of clinical trials results of MRD-testing should be used to direct CLL therapy. There is a consensus the answer is no.

MRD AS A SURROGATE ENDPOINT FOR DRUG DEVELOPMENT AND REGULATORY APPROVAL

From a regulatory perspective, drug approval requires proving safety and efficacy. In CLL PFS is often the endpoint used because of the long interval required to evaluate survival and confounding by subsequent interventions such that any RCT eventuates in an observational database. However, as we discuss above, although PFS is a reasonable endpoint for drug approvals, CIR is the biologically relevant endpoint for a test designed to quantify numbers of residual leukaemia cells, biologically relevant or not and within a given interval or not. In so far as a regulatory agency is interested in a clinical surrogate endpoint using MRD-test results for approval seems reasonable and is the position seemingly adopted by the European Medicines Agency (EMA). However, this position is not currently shared by US FDA which recommends

MRD-testing only in persons with a complete response [59]. Reasons why US FDA might take this position, are discussed above. We also note that in many if not most cancers PFS has proved an unreliable survival surrogate [60, 61]. So is using results of MRD-testing for drug approvals reasonable? Perhaps; but only perhaps. The jury is out whether the US FDA or EMA opinion is the winner.

CONCLUSION

Advances using diverse techniques have resulted in MRD-assays with reasonably high sensitivity and specificity but there are issues with false-positive and -negative test results, PPV and NPV and with reproducibility and replicability. It's also unclear how to interpret discordant MRD-test results in someone, a situation that may be intractable for reasons we discuss. Because of these considerations result of MRD-testing should not be used to direct therapy of persons with CLL. MRD-testing is being more widely used but standardization and harmonization are needed to increase precision and enable comparisons between studies and to determine accuracy and precision of MRD-testing in predicting relapse hazard and as a surrogate endpoint for drug approvals. Work toward this goal by ERIC is ongoing.

Despite the limitations we discuss, many studies report results of MRD-testing correlates with PFS after most CLL therapies in cohorts of persons with CLL whether or not they achieve a iwCLL-defined complete response. This raises the question of whether the iwCLL definition of complete response should be abandoned as an efficacy endpoint. We think not, or at least not yet.

At the individual level, results of MRD-testing at one landmark time point may only slightly increase the prediction compared with that currently achieved. The possibility of false-positive and -negative MRD-tests and sampling error needs consideration. To what degree prediction accuracy increases with repeated MRD-testing is unknown but seems likely. Whether and how one should respond to results of data from MRD-testing remains to be determined by appropriately designed clinical trials. An important variable is physician and patient tolerance for an incorrect MRD-test result and consequences of acting thereon. The treating physician must evaluate data from cohorts of similar persons with CLL combined with an understanding of the limitations of the MRD- to be able to estimate outcomes and advise patients appropriately. For example, if the therapy prompted by the MRD-test result has little or no risk of adverse events one might be willing to accept a false-positive MRD-test. However, if the consequences could be dire, tolerance for a false-positive MRD-test should be less. The treating physician must evaluate data from cohorts of similar persons together with an understanding of the limitations of the MRD-testing and with a person-specific objective and subjective co-variables to gain informed consent.

Conventionally defined criteria of complete response in CLL encompass widely diverse levels of residual leukaemia cells and correlates with diverse clinical outcomes. Consequently, we believe some estimation of MRD by *state-of-the-art* MRD-testing (typically ≥ 2 metachronous time points with orthogonal testing) is now and in the future incorporated into CLL clinical trials. Costs of integrating MRD-testing into clinical trials and the lack of a universally accepted MRD assay are challenging but need not be prohibitive. RCTs comparing different MRD-tests in heterogeneous populations of persons with CLL at diverse times during therapy and across different therapies are needed.

Data from clinical trials could potentially prove useful if carefully annotated with details of MRD-test performance. The importance of randomized assessments of the impact of acting on the results of an MRD-test is clear. We need to know whether a positive MRD-test accurately identifies persons with a poor prognosis and, if so, convincing evidence an intervention(s) can alter this. We see an interesting and potentially bright future for MRD-testing in CLL

but are reminded of the wisdom in the adage from Alexander Pope: Be not the first by whom the new are tried, Nor yet the last to lay the old aside and of the Dunning–Kruger effect of which we pray we are not guilty [62, 63].

Shenmiao Yang¹ , Neil E. Kay² , Min Shi³ ,
Gert Ossenkoppele⁴, Roland B. Walter⁵ and
Robert Peter Gale⁶

¹Peking University Peoples Hospital, Peking University Institute of Hematology, Beijing, China. ²Mayo Clinic, Department of Medicine, Division of Hematology, Rochester, MN, USA. ³Mayo Clinic, Department of Laboratory Medicine and Pathology, Rochester, MN, USA. ⁴Department of Hematology, Amsterdam UMC, Vrije Universiteit Amsterdam, Cancer Center Amsterdam, Amsterdam, The Netherlands. ⁵Fred Hutchinson Cancer Research Center, Clinical Research Division, Seattle, WA, USA. ⁶Haematology Research Centre, Department of Immunology and Inflammation, Imperial College London, London, UK. ✉email: robertpetergale@alumni.ucla.edu

REFERENCES

- Algrin C, Golmard JL, Michallet M, Reman O, Huynh A, Perrot A, et al. Flow cytometry minimal residual disease after allogeneic transplant for chronic lymphocytic leukemia. *Eur J Haematol*. 2017;98:363–70.
- Thompson PA, Tam CS, O'Brien SM, Wierda WG, Stingo F, Plunkett W, et al. Fludarabine, cyclophosphamide, and rituximab treatment achieves long-term disease-free survival in IGHV-mutated chronic lymphocytic leukemia. *Blood*. 2016;127:303–9.
- Uchiyama T, Yokoyama A, Aoki S. Measurable residual disease in the treatment of chronic lymphocytic leukemia. *J Clin Exp Hematop*. 2020;60:138–45.
- Heltai S, Ghia P, Scarfo L. Relevance of minimal residual disease in the era of targeted agents. *Cancer J*. 2019;25:410–7.
- Tomuleasa C, Selicean C, Cismas S, Jurj A, Marian M, Dima D, et al. Minimal residual disease in chronic lymphocytic leukemia: a consensus paper that presents the clinical impact of the presently available laboratory approaches. *Crit Rev Clin Lab Sci*. 2018;55:329–45.
- Owen C, Christofides A, Johnson N, Lawrence T, MacDonald D, Ward C. Use of minimal residual disease assessment in the treatment of chronic lymphocytic leukemia. *Leuk Lymphoma*. 2017;58:2777–85.
- Thompson PA, Wierda WG. Eliminating minimal residual disease as a therapeutic end point: working toward cure for patients with CLL. *Blood*. 2016;127:279–86.
- Botthcher S, Hallek M, Ritgen M, Kneba M. The role of minimal residual disease measurements in the therapy for CLL: is it ready for prime time? *Hematol Oncol Clin North Am*. 2013;27:267–88.
- Varghese AM, Rawstron AC, Hillmen P. Eradicating minimal residual disease in chronic lymphocytic leukemia: should this be the goal of treatment? *Curr Hematol Malig Rep*. 2010;5:35–44.
- Moreno C, Ritgen M, Rawstron A. Is MRD eradication a desirable goal in CLL? *Best Pr Res Clin Haematol*. 2010;23:97–107.
- Harrison WJ. The total cellularity of the bone marrow in man. *J Clin Pathol*. 1962;15:254–9.
- de The H, Lavau C, Marchio A, Chomienne C, Degos L, Dejean A. The PML-RAR alpha fusion mRNA generated by the t(15;17) translocation in acute promyelocytic leukemia encodes a functionally altered RAR. *Cell*. 1991;66:675–84.
- Goldman JM, Melo JV. Chronic myeloid leukemia—advances in biology and new approaches to treatment. *N Engl J Med*. 2003;349:1451–64.
- Shtivelman E, Lifshitz B, Gale RP, Canaani E. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. *Nature*. 1985;315:550–4.
- Saussele S, Richter J, Guilhot J, Gruber FX, Hjorth-Hansen H, Almeida A, et al. Discontinuation of tyrosine kinase inhibitor therapy in chronic myeloid leukaemia (EURO-SKI): a prespecified interim analysis of a prospective, multicentre, non-randomised, trial. *Lancet Oncol*. 2018;19:747–57.
- Hochhaus A, Baccarani M, Silver RT, Schiffer C, Apperley JF, Cervantes F, et al. European LeukemiaNet 2020 recommendations for treating chronic myeloid leukemia. *Leukemia*. 2020;34:966–84.
- Hoffmann VS, Baccarani M, Hasford J, Castagnetti F, Di Raimondo F, Casado LF, et al. Treatment and outcome of 2904 CML patients from the EUTOS population-based registry. *Leukemia*. 2017;31:593–601.
- Hochhaus A, Weisser A, La Rosee P, Emig M, Muller MC, Saussele S, et al. Detection and quantification of residual disease in chronic myelogenous leukemia. *Leukemia*. 2000;14:998–1005.
- van der Velden VH, Hochhaus A, Cazzaniga G, Szczepanski T, Gabert J, van Dongen JJ. Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia*. 2003;17:1013–34.
- Houshmand M, Simonetti G, Circosta P, Gaidano V, Cignetti A, Martinelli G, et al. Chronic myeloid leukemia stem cells. *Leukemia*. 2019;33:1543–56.
- Mahon FX, Rea D, Guilhot J, Guilhot F, Huguet F, Nicolini F, et al. Discontinuation of imatinib in patients with chronic myeloid leukaemia who have maintained complete molecular remission for at least 2 years: the prospective, multicentre Stop Imatinib (STIM) trial. *Lancet Oncol*. 2010;11:1029–35.
- Butturini A, Klein J, Gale RP. Modeling minimal residual disease (MRD)-testing. *Leuk Res*. 2003;27:293–300.
- Botthcher S, Ritgen M, Fischer K, Stilgenbauer S, Busch RM, Fingerle-Rowson G, et al. Minimal residual disease quantification is an independent predictor of progression-free and overall survival in chronic lymphocytic leukemia: a multivariate analysis from the randomized GCLLSG CLL8 trial. *J Clin Oncol*. 2012;30:980–8.
- Goede V, Fischer K, Busch R, Engelke A, Eichhorst B, Wendtner CM, et al. Obinutuzumab plus chlorambucil in patients with CLL and coexisting conditions. *N Engl J Med*. 2014;370:1101–10.
- Eichhorst B, Fink AM, Bahlo J, Busch R, Kovacs G, Maurer C, et al. First-line chemoimmunotherapy with bendamustine and rituximab versus fludarabine, cyclophosphamide, and rituximab in patients with advanced chronic lymphocytic leukaemia (CLL10): an international, open-label, randomised, phase 3, non-inferiority trial. *Lancet Oncol*. 2016;17:928–42.
- Thompson PA. MRD negativity as a surrogate for PFS in CLL? *Blood*. 2018;131:943–4.
- Rawstron AC, Fazi C, Agathangelidis A, Villamor N, Letestu R, Nomdedeu J, et al. A complementary role of multiparameter flow cytometry and high-throughput sequencing for minimal residual disease detection in chronic lymphocytic leukemia: an European Research Initiative on CLL study. *Leukemia*. 2016;30:929–36.
- Thompson PA, Srivastava J, Peterson C, Strati P, Jorgensen JL, Hether T, et al. Minimal residual disease undetectable by next-generation sequencing predicts improved outcome in CLL after chemoimmunotherapy. *Blood*. 2019;134:1951–9.
- Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*. 2014;371:2488–98.
- Fabbri G, Dalla-Favera R. The molecular pathogenesis of chronic lymphocytic leukaemia. *Nat Rev Cancer*. 2016;16:145–62.
- Sutton LA, Ljungstrom V, Enjuanes A, Cortese D, Skafstrom A, Tausch E, et al. Comparative analysis of targeted next-generation sequencing panels for the detection of gene mutations in chronic lymphocytic leukemia: an ERIC multicenter study. *Haematologica*. 2021;106:682–91.
- Malcikova J, Tausch E, Rossi D, Sutton LA, Soussi T, Zenz T, et al. ERIC recommendations for TP53 mutation analysis in chronic lymphocytic leukemia—update on methodological approaches and results interpretation. *Leukemia*. 2018;32:1070–80.
- Al-Sawaf O, Zhang C, Tandon M, Sinha A, Fink AM, Robrecht S, et al. Venetoclax plus obinutuzumab versus chlorambucil plus obinutuzumab for previously untreated chronic lymphocytic leukaemia (CLL14): follow-up results from a multicentre, open-label, randomised, phase 3 trial. *Lancet Oncol*. 2020;21:1188–200.
- Al-Sawaf O, Zhang C, Robrecht S, Tandon M, Panchal A, Fink A, et al. Venetoclax-obinutuzumab for previously untreated chronic lymphocytic leukemia: 4-year follow-up analysis of the randomized CLL14 study. Paper presented at: European Hematology Association 2021 Virtual Congress; June 2021; Abstract S146. 2021; 2021.
- Huang H-Y, Wang Y, Zhang Y-C, Tobias H, Gale RP, Li J-Y, et al. A prognostic survival model of the immune microenvironment in persons with chronic lymphocytic leukemia.
- Kater AP, Kipps TJ, Eichhorst B, Hillmen P, D'Rozario J, Owen C, et al. Five-year analysis of Murano study demonstrates enduring undetectable minimal residual disease (uMRD) in a subset of relapsed/refractory chronic lymphocytic leukemia (R/R CLL) patients (Pts) following fixed-duration venetoclax-rituximab (VenR) therapy (Tx). *Blood*. 2020;136:19–21.
- Ommen HB, Schnitger S, Jovanovic JV, Ommen IB, Hasle H, Ostergaard M, et al. Strikingly different molecular relapse kinetics in NPM1c, PML-RARA, RUNX1-RUNX1T1, and CBFB-MYH11 acute myeloid leukemias. *Blood*. 2010;115:198–205.
- Hokland P, Ommen HB. Towards individualized follow-up in adult acute myeloid leukemia in remission. *Blood*. 2011;117:2577–84.
- Lo Coco F, Diverio D, Avvisati G, Petti MC, Meloni G, Pogliani EM, et al. Therapy of molecular relapse in acute promyelocytic leukemia. *Blood*. 1999;94:2225–9.
- Jain N, Thompson P, Burger J, Ferrajoli A, Takahashi K, Estrov Z, et al. Ibrutinib, fludarabine, cyclophosphamide, and obinutuzumab (iFCG) regimen for chronic lymphocytic leukemia (CLL) with mutated IGHV and without TP53 aberrations. *Leukemia* 2021; 2021, <https://doi.org/10.1038/s41375-021-01280-8>.

41. Ni W, Hu B, Zheng C, Tong Y, Wang L, Li QQ, et al. Automated analysis of acute myeloid leukemia minimal residual disease using a support vector machine. *Oncotarget*. 2016;7:71915–21.
42. Langerak AW, Ritgen M, Goede V, Robrecht S, Bahlo J, Fischer K, et al. Prognostic value of MRD in CLL patients with comorbidities receiving chlorambucil plus obinutuzumab or rituximab. *Blood*. 2019;133:494–7.
43. Kramer I, Stilgenbauer S, Dietrich S, Bottcher S, Zeis M, Stadler M, et al. Allogeneic hematopoietic cell transplantation for high-risk CLL: 10-year follow-up of the GCLLSG CLL3X trial. *Blood*. 2017;130:1477–80.
44. Tam CS, Siddiqi T, Allan JN, Kipps TJ, Flinn IW, Kuss BJ, et al. Ibrutinib (Ibr) plus venetoclax (Ven) for first-line treatment of chronic lymphocytic leukemia (CLL)/Small lymphocytic lymphoma (SLL): results from the MRD cohort of the phase 2 CAPTIVATE study. *Blood*. 2019;134:35.
45. Hourigan CS, Gale RP, Gormley NJ, Ossenkoppele GJ, Walter RB. Measurable residual disease testing in acute myeloid leukaemia. *Leukemia*. 2017;31:1482–90.
46. Othus M, Gale RP, Hourigan CS, Walter RB. Statistics and measurable residual disease (MRD) testing: uses and abuses in hematopoietic cell transplantation. *Bone Marrow Transpl*. 2020;55:843–50.
47. Kater AP, Seymour JF, Hillmen P, Eichhorst B, Langerak AW, Owen C, et al. Fixed duration of venetoclax-rituximab in relapsed/refractory chronic lymphocytic leukemia eradicates minimal residual disease and prolongs survival: post-treatment follow-up of the MURANO phase III study. *J Clin Oncol*. 2019;37:269–77.
48. Hourigan CS, Dillon LW, Gui G, Logan BR, Fei M, Ghannam J, et al. Impact of conditioning intensity of allogeneic transplantation for acute myeloid leukemia with genomic evidence of residual disease. *J Clin Oncol*. 2020;38:1273–83.
49. Craddock C, Jackson A, Loke J, Siddique S, Hodgkinson A, Mason J, et al. Augmented reduced-intensity regimen does not improve post-allogeneic transplant outcomes in acute myeloid leukemia. *J Clin Oncol*. 2021;39:768–78.
50. Yeoh AE, Ariffin H, Chai EL, Kwok CS, Chan YH, Ponnudurai K, et al. Minimal residual disease-guided treatment deintensification for children with acute lymphoblastic leukemia: results from the Malaysia-Singapore acute lymphoblastic leukemia 2003 study. *J Clin Oncol*. 2012;30:2384–92.
51. Eckert C, Henze G, Seeger K, Hagedorn N, Mann G, Panzer-Grümayer R, et al. Use of allogeneic hematopoietic stem-cell transplantation based on minimal residual disease response improves outcomes for children with relapsed acute lymphoblastic leukemia in the intermediate-risk group. *J Clin Oncol*. 2013;31:2736–42.
52. Roberts KG, Pei D, Campana D, Payne-Turner D, Li Y, Cheng C, et al. Outcomes of children with BCR-ABL1-like acute lymphoblastic leukemia treated with risk-directed therapy based on the levels of minimal residual disease. *J Clin Oncol*. 2014;32:3012–20.
53. Pui CH, Pei D, Coustan-Smith E, Jeha S, Cheng C, Bowman WP, et al. Clinical utility of sequential minimal residual disease measurements in the context of risk-based therapy in childhood acute lymphoblastic leukaemia: a prospective study. *Lancet Oncol*. 2015;16:465–74.
54. Vora A, Goulden N, Wade R, Mitchell C, Hancock J, Hough R, et al. Treatment reduction for children and young adults with low-risk acute lymphoblastic leukaemia defined by minimal residual disease (UKALL 2003): a randomised controlled trial. *Lancet Oncol*. 2013;14:199–209.
55. Vora A, Goulden N, Mitchell C, Hancock J, Hough R, Rowntree C, et al. Augmented post-remission therapy for a minimal residual disease-defined high-risk subgroup of children and young people with clinical standard-risk and intermediate-risk acute lymphoblastic leukaemia (UKALL 2003): a randomised controlled trial. *Lancet Oncol*. 2014;15:809–18.
56. Schrappe M, Bleckmann K, Zimmermann M, Biondi A, Moricke A, Locatelli F, et al. Reduced-intensity delayed intensification in standard-risk pediatric acute lymphoblastic leukemia defined by undetectable minimal residual disease: results of an international randomized trial (AIEOP-BFM ALL 2000). *J Clin Oncol*. 2018;36:244–53.
57. Esteve J, Escoda L, Martín G, Rubio V, Díaz-Mediavilla J, González M, et al. Outcome of patients with acute promyelocytic leukemia failing to front-line treatment with all-trans retinoic acid and anthracycline-based chemotherapy (PETHEMA protocols LPA96 and LPA99): benefit of an early intervention. *Leukemia*. 2007;21:446–52.
58. Grimwade D, Jovanovic JV, Hills RK. Can we say farewell to monitoring minimal residual disease in acute promyelocytic leukaemia? *Best Pr Res Clin Haematol*. 2014;27:53–61.
59. Center for drug evaluation and research. center for biologics evaluation and research. hematologic malignancies: regulatory considerations for use of minimal residual disease in development of drug and biological products for treatment guidance for industry January 2020. Docket Number: FDA-2018-D-3090. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/hematologic-malignancies-regulatory-considerations-use-minimal-residual-disease-development-drug-and>.
60. Kim C, Prasad V. Cancer drugs approved on the basis of a surrogate end point and subsequent overall survival: an analysis of 5 years of US food and drug administration approvals. *JAMA Intern Med*. 2015;175:1992–4.
61. Belin L, Tan A, De Rycke Y, Dechartres A. Progression-free survival as a surrogate for overall survival in oncology trials: a methodological systematic review. *Br J Cancer*. 2020;122:1707–14.
62. Pope A. An Essay on Criticism. London 1711 in The poems of Alexander Pope: a one-volume edition of the Twickenham text with selected annotations: Yale Univ. Press, New have; 1963.
63. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol*. 1999;77:1121–34.

ACKNOWLEDGEMENTS

Profs. Christopher Hourigan (US National Institutes of Health), Emli Monserrat (Univ. Barcelona) and Robin Foa (Università “La Sapienza” di Roma) kindly reviewed the typescript and made helpful suggestions. RPG acknowledges support from the National Institute of Health Research (NIHR) Biomedical Research Centre funding scheme.

AUTHOR CONTRIBUTIONS

SY and RPG conceived the typescript. They and KN and MS developed the 1st draft. RW and GO made substantial contributions. All authors contributed to the final typescript, are responsible for the content and agreed to publication.

COMPETING INTERESTS

NEK is on the Advisory Board for Abbvie, Astra Zeneca, Behring, Cytomx Therapy, Dava Oncology, Janssen, Juno Therapeutics, Oncotracker, Pharmacyclics and Targeted Oncology. DSMC (Data Safety Monitoring Committee) for Agios Pharm, Astra Zeneca, BMS –Celgene, Cytomx Therapeutics, Janssen, Morpho-sys, Rigel. He received research funding from Abbvie, Acerta Pharma, Bristol Meyer Squib, Celgene, Genentech, MEI Pharma, Pharmacyclics, Sunesis, TG Therapeutics and Tolero Pharmaceuticals. RPG is a consultant to BeiGene Ltd., Fusion Pharma LLC, LaJolla NanoMedical Inc., Mingsight Pharmaceuticals Inc., CStone Pharmaceuticals, NexImmune Inc. and Prolacta Bioscience; advisor to Antengene Biotech LLC, Medical Director, FFF Enterprises Inc.; Partner, AZCA Inc.; Board of Directors, Russian Foundation for Cancer Research Support; and Scientific Advisory Board: StemRad Ltd.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Robert Peter Gale.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.