Quantifying mutational synergy using computational models predicts survival in haematological cancers

Richard Norris [1], John Jones[1], Erika Mancini[2], Timothy Chevassut[1], Chris Pepper[1], Andrea Pepper[1], Simon Mitchell[1*]

[1]Brighton and Sussex Medical School, Department of Clinical and Experimental Medicine, Brighton, United Kingdom

[2]School of Life Sciences, University of Sussex, Brighton, United Kingdom

* Correspondence: S.A.Mitchell@bsms.ac.uk

Running Title: Mutational synergy predicts survival

Abstract

Genetic heterogeneity and co-occurring driver mutations contribute to poor clinical outcomes in cancer. However, the impact of multiple mutations on complex signalling networks is not easily predicted. We found that, by placing mutations into their cellular context, multi-scale agent-based mathematical models could predict how genetic events combine in haematological malignancies. Simulations of lymphoma and myeloma predicted co-occurring mutations synergised to increase tumour cell expansion beyond what would be expected from the impact of the individual mutations alone. Mutational synergy between *MYC* and *BCL2* was consistent with the more aggressive disease course of patients with double-hit lymphoma, and mutational synergy between MCL1 and CKS1B was predictive of outcome in patients with gain 1q multiple myeloma. Incorporating patient-specific mutational profiles into personalised models of lymphoma patients with the worst clinical outcomes revealed a correlation between simulated mutational synergy and overall survival, which outperformed widely used classifications of lymphoma informed by gene expression or mutational data alone. Our results demonstrated that mutational synergy scores enabled prediction of the impact of co-occurring mutations and may improve personalised prognostic predictions.

Keywords: Cancer / Computational Biology / DLBCL / Myeloma / Systems Biology

Introduction

Mutational heterogeneity represents a major barrier to understanding the molecular mechanisms responsible for tumour progression, and to the development of novel treatment strategies. With the advent of whole exome sequencing (WES) many malignancies, including B-cell malignancies such as Diffuse Large B-Cell Lymphoma (DLBCL) and Multiple Myeloma (MM), have become characterised by interpatient mutational heterogeneity and within-patient mutational complexity (1, 2). This factor contributes to variation in response to current treatments such that first line therapy for DLBCL is ineffective for around 40% of patients, while almost all MM patients will eventually relapse. The most aggressive malignancies often contain genetic aberrations affecting multiple genes. 'Double hit' (DH) DLBCL, featuring overexpression of *MYC* and *BCL2* (or *BCL6*), is among the most aggressive lymphoid malignancies, with very poor outcomes for these patients (3, 4). The prognostic implications of DH lymphomas were the basis for the 2017 update on DLBCL by the World Health Organization. Increased expression of *MYC*, which codes for the transcription factor Myc (also known as c-Myc), is a significant factor in the growth of multiple tumour types (5), and mutations affecting *MYC* occur in approximately 33% of DLBCL patients (6). Aberrant expression of *BCL2* is anti-apoptotic (7), and as such the genes defining DH lymphoma represent archetypal examples of two of the hallmarks of cancer: sustained proliferation and resistance to cell death (8). *BCL6* encodes a transcriptional repressor expressed in healthy germinal centre B-cells. Like *BCL2*, over-expression of *BCL6* can prevent apoptosis in DLBCL (9).

Initial attempts to characterise the heterogeneity of DLBCL utilised gene expression profiling that split cases into subgroups based on their putative cell of origin (germinal centre – GC, or activated B cell – ABC) (10). Subsequent studies leveraged WES to identify 5 or more

prognostically informative clusters of patients based on their mutational profiles (11-13). LymphGen identified a DH gene expression signature in the *EZB-MYC*$^{+}$ genetic cluster, characterised by *BCL2* translocation in 72% of cases, but only bearing genetic abnormalities in *MYC* in 38% of cases (13). Double hit gene expression signatures have also been identified in tumours lacking rearrangements of *MYC* and/or *BCL2* (14). Overall, fewer than 10% of DLBCL cases are true DH lymphoma, yet 30-40% of DLBCL patients relapse following frontline treatment and have a poor prognosis (15). Comparisons can be made in MM, where specific genetic events are known to confer a more aggressive phenotype. An example is gain of chromosome 1q, which occurs in 30-50% of tumours and is associated with poorer outcome (16, 17). However, even in the absence of known high-risk genetic changes, around 70% of patients will succumb to the disease within 10 years. Taken together these data demonstrate limits to our ability to identify combinations of mutations that confer aggressive disease and poor prognosis. Importantly, clustering patients by mutational burden ignores the context of the complex gene regulatory signalling networks involving mutated genes and the gene products they encode.

Computational models of B-cell fates have provided insight into healthy immune responses, with predictions of cell proliferation and survival validated by laboratory experiments (18-21). Here, we used mechanistic models to simulate how mutations combine in B-cell malignancies. We quantified the emergent behaviour resulting from combinations of mutations impacting multiple signalling networks by simulating individual mutations and co-occurring mutations in the same cell. We developed a novel method to quantify the extent to which mutations combine to confer a worse prognosis than expected based on the impact of each individual mutation in isolation. Using WES data to create individual virtual DLBCL patients, we found that quantification of mutational synergy predicted overall survival better than clustering (of WES data) or gene expression-based classification. To further validate our model in an alternative B cell malignancy

we modelled the impact of gain(1q) MM. Chromosome 1q is commonly amplified in MM, but the impact of putative driver genes in this chromosomal region on prognosis is not well understood. We computationally identify a gene-dosage effect, with increasing copies of chromosome 1q correlating with increased mutational synergy, a prediction that aligns with outcomes in clinical trials. Taken together, we demonstrated that “multi-hit” haematological malignancies, with poor prognosis resulting from co-occurring driver mutations, can be identified by incorporating mutational data into systems biology modelling to quantify mutational synergy.

Results

Simulating individual mutations is insufficient to explain aggressive disease.

Computational systems biology provides a unique opportunity to predict the impact of mutations on individual molecular networks in isolation, as it is challenging to isolate the molecular machinery controlling proliferation and apoptosis in cell-based systems. Initially, we sought to model the effect of gain-of-function mutations in *MYC* and *BCL2* on cell cycle and apoptosis signalling pathways, prior to investigating the effect of combining the mutations. Given the broad impact of c-Myc, modelling its actions is challenging. However, previous modelling work was able to recapitulate both single cell and bulk experimental measurements by encoding c-Myc as acting on B-cell growth in cell cycle models (18, 19). Modelling the function of the Bcl2 family of proteins meanwhile was centred around programmed cell death, and Bcl2 itself an inhibitor of apoptosis (22).

Simulating the impact of overexpression of *MYC* predicted a modest impact on the timing of cell cycle phases and overall time to divide (Fig 1A). A sensitivity analysis also predicted that increased *MYC* expression has a small effect on cell cycle time relative to expression changes of other genes in the pathway (Fig S1). This result recapitulated experimental tracking of

division times in B-cells transfected with c-Myc and cellular proliferation measurement in lymphocytes from Eμ-Myc mice (23). In simulations of increasing levels of *MYC* expression, the decrease in total cell cycle time was primarily due to a shortening of G1 phase. A dose-dependent effect was seen with increasing *MYC* expression up to the equivalent of 5 extra copies, after which no further effect is seen (Fig 1A, Fig S2). Simulating the impact of *BCL2* overexpression, in a computational model of apoptosis (24), predicted a dose-dependent delay of poly ADP-ribose polymerase (PARP) cleavage up to the equivalent of 5 extra copies of *BCL2*, beyond which the cells survive the full length of the simulation (Fig 1B). These simulations represented an average B-cell using published parameters, but it is not clear that mutations will affect each cell equally within a heterogeneous population of individual B-cells.

Simulating a population of molecularly heterogeneous individual B-cells, by incorporating cell-to-cell heterogeneity as described previously (18), resulted in cell cycle timings that recapitulated the range and distribution of cell cycle duration and cell death timings measured by time-lapse microscopy in murine B lymphocytes (25) (Fig 1C and D). Increased *MYC* expression was predicted to have a minor effect (decreasing total cell cycle time) in all cells (Fig 1C). These simulations also predicted a small population of cells (~80/1000) to be in cell cycle arrest in the WT simulation. Simulated cells in this state did not progress towards a constant cell cycle time (limit cycle), but rather their cell mass fell due to attempts to proliferate faster than they could gain mass, which lead to cell cycle arrest. Concentrations of the cell cycle inhibitor p27, and complexes containing p27, were predicted to be upregulated in these arrested cells (Fig S3), which corresponds with higher p27 levels in experimental measurements of cell cycle arrest (24). *MYC* overexpression was predicted to reverse the cell cycle arrest in >80% of these cells, and strikingly they become the most rapidly proliferating cells (Fig 1C). The number of cells switched was proportional to the level of *MYC* expression (Fig S4). *MYC* is amongst the most differentially expressed genes in quiescence and reverses quiescence in haematopoietic stem

cells (26, 27), therefore it is plausible that this effect may also be present in B-cells. The impact of *BCL2* over-expression was predicted to proportionally increase with the survival time of the wild-type cell, with the largest Impact in cells molecularly predisposed for long survival (Fig 1E and F).

As the fates of cell division and apoptosis are mutually exclusive and competing fates in individual cells (dead cells can't divide, while cells that survive may divide again) (18, 28), how these two fates combine in individual cells with DH lymphoma is unclear. To gain a better understanding of this, simulating these signalling networks simultaneously in individual cells is required. If within a DH cell population there exists individual cells that concurrently have the shortest cell cycle times (Fig 1C) and the longest survival times (Fig 1E), then the mutations are likely to combine deleteriously. However, if the most rapidly cycling cells are also the most rapidly dying cells, then the combined effect of the mutations is unlikely to be greater than expected from their individual impact.

Multi-scale modelling enables quantification of mutational synergy.

To predict how the DH mutations combine in single cells, we employed multi-scale, agent-based models of B-cells, which enable simulation of multiple signalling networks simultaneously in a heterogeneous population of individual cells. We used an established multi-scale model of B-cell fates that enables simulation of proliferation, apoptosis and terminal B-cell differentiation (Fig 2A). The input to this simulation was mitogenic signalling through NF-κB essential modulator (NEMO)-IκB kinase (IKK) control over the Nuclear Factor κB (NF-κB) transcription factor. In this modelling framework NF-κB induced pro-proliferation and anti-apoptotic target genes within the cell cycle and apoptosis modules of the model. In this way we could simulate a cell population in which cells that undergo apoptosis were removed from the cell population

(determined by an increase in cleaved PARP), while cells that divide introduced new simulations of daughter cells (determined by an increase in Cdh1, which increases upon completion of mitosis) (Fig 2A). These models have been validated against cell-population and single-cell data from primary B-ells, and allowed us to simulate large heterogeneous populations of virtual B-cells (18, 19). We sought to use this modelling approach to quantify the predicted effect of mutations impacting multiple signalling pathways. We developed a method of visualising and quantifying mutational synergy to evaluate the degree to which multiple simulated mutations confer increased cell population size compared to what would be expected from the sum of the effects of the individual mutations (Fig 2B). Here, synergy was assumed to be deleterious, as more synergy results in more proliferative disease.

Mutational synergy in simulations aligns with poor prognosis in patients.

The results of simulations of individual mutations in individual cells (Fig 1) indicated that overexpression of *MYC* would result in a small increase in cell population size, while the overexpression of *BCL2* would have a larger effect, and this was indeed seen in multi-scale modelling results (Fig 2C, red region). Interestingly, the effect of *BCL6* and *MYC* overexpression on cell number is predicted to be similar, but less than *BCL2* overexpression (Fig 2D). We found that the *MYC*+*BCL2* double hit simulations, with each gene upregulated 1.5-fold (approximating one additional copy of each gene), showed a small amount of synergy, producing larger cell populations than the additive effects of individual mutations (Fig 2C, red region). This synergistic effect was predicted to be stronger for the DH featuring *MYC*+*BCL2* than for *MYC*+*BCL6* (Fig 2C and 2D). This difference was found to be more pronounced when DH genes were overexpressed 5 times (vs x1.5), which is consistent with the 4-6 fold change in gene expression seen in lymphoma patient samples with genomic events affecting *BCL2* and *MYC* (29)(Fig S5 and S6). Interestingly, this difference in effect size on mutations synergy when

MYC+BCL6 are overexpressed x1.5 to x5 is under half the size seen with increased expression of *MYC+BCL2* (Fig 2F compared to Fig 2E and Fig S5). If mutational synergy correlates with prognosis, these simulations predicted that double hits involving *BCL2* are more likely to confer aggressive disease than those involving *BCL6* (Fig 2C-F). Indeed, this was consistent with results from a multi-centre study that found the prognosis for *MYC+BCL2* double hits was significantly worse than for *MYC+BCL6* (30). The dose-dependent increase in synergy with extra copies of *BCL2* and *MYC* also aligns with clinical outcome as *MYC* rearrangements, in which the gene is highly expressed, confer worse prognosis than copy number alterations (31).

Given the alignment between mutational synergy and outcomes we found for lymphoma, and to establish whether the utility of simulated mutational synergy is restricted to lymphoma, we simulated the impact of genetic events that commonly occur in cases of MM with poor prognosis. Gain 1q MM is a commonly occurring cytogenetic event in MM associated with therapeutic failure and shortened survival (32). Simulating chromosomal gains with the multi-scale modelling framework required mapping genes on the chromosome to model parameters. Genes encoding the Bcl2-family protein Mcl1 and cell cycle regulator Cks1b both reside on chromosome 1q21 and have been implicated in the pathogenesis of gain1q MM (33). We mapped these genes to modelled parameters (see methods) and found a similar pattern to DH lymphoma, i.e., a small amount of synergy for x1.5-fold increase (equivalent to one extra copy) and increasing synergy proportional to chromosome 1q copy number (Fig 3A-C). Strikingly, with 5-fold increases in expression of *MCL1* and *CKS1b* the synergistic effect is such that the cell population continues to increase in size throughout the simulation, even beyond the point at which the input signal (IKK activity) is removed (Fig 3C, and IKK input curve; Fig 2A). The dose-dependent increase in mutational synergy aligns with MM patient outcome data, which shows that prognosis worsens with increasing copies of chromosome 1q (Fig 3D and Fig S7). Taken

together we found that by mapping mutational events and chromosomal abnormalities to model parameters, then simulating the impact of individual mutations and the combined impact of multiple mutations, the resultant mutational synergy correlated with outcomes in patients with poor prognosis lymphoma and MM.

Disease severity can be predicted from multi-hit mutational synergy in patients with poor prognosis.

Only a small subgroup of lymphoma patients (~5%) have well-defined DH lymphoma. Despite this, over 20% of patients will not survive 3 years after diagnosis. While co-morbidities and non-genetic changes likely contribute to this variability in prognosis, we wondered whether substantially more patients exist with complex mutational combinations that confer poor prognosis. Given the observations from simulating DH lymphoma and MM we hypothesised that calculating per-patient mutational synergy using the mutational profile of individual patients may reveal a correlation with overall survival. To test this hypothesis, we leveraged comprehensive WES data from patients with long-term follow-up (Fig 4A)(11). To create virtual patients, we mapped mutated genes and chromosomal abnormalities to modelled parameters (see Methods). Initial analysis, in which we visualised peak cell population size and mutational synergy in a small subset of patients, indicated that patients predicted to have high mutational synergy had relatively poor prognoses (Fig 4B & 4C).

To test whether simulated mutational synergy can predict overall survival we simulated both individual mutations and the combined mutational profile of all deceased DLBCL patients in the WES cohort with OS <36 months (11); the majority of events in trials that stratify DH lymphoma patients occur in the first 3 years (11, 34, 35). This analysis required simulating 100 founder cells for each mutation which, following cell proliferation within the agent-based modelling framework, resulted in simulating 100-893 cells per condition. Performing this analysis for these

patients required solving 15783 individual cells, and a total of 306902 ordinary differential equations. This analysis resulted in simulated cell numbers for each mutation reported in the WES data for this cohort, and cell numbers for the combination of mutations found in each patient. We assumed the effect size of each mutation was 1.5-fold for upregulation or 0.5 for downregulation. Simulations for every patient are provided in the supplementary material. To quantify the magnitude of synergy we created a mutational synergy score (MSS), defined as the difference in cell number between the sum of individual mutations and combinations of all mutations, normalised to the cell population size for each patient. MSS can be visualised as the largest height of the red region in Fig 2, 3 and 4B-C, normalised to the peak red line in each plot.

Comparing MSS to OS revealed a significant inverse correlation between the MSS and survival (Fig 4D R^2 = 21.2%, p = 0.0156 using a generalised linear regression analysis). Simulated cell number alone (peak simulated cell number rather than MSS) did not correlate as strongly with OS (R^2=0.1, Fig S8). Interestingly, when we compared MMS in the DLBCL cohort to those calculated for the classic DH simulations, we identified two patients with higher synergy scores than we found for 5-fold overexpression of both *MYC* and *BCL2* 5x (Fig 2E, Fig S6, MSS= 0.76). Both patients have OS <7 months and co-occurring mutations impacting NF-κB (REL copy number increase), apoptosis (*BCL2* or *MCL1*), and blocking B-cell differentiation (*BLIMP1* loss). While one patient also has a cell cycle mutation (*RB1*), the other lacks cell cycle activating mutations but may be driven into a highly proliferative regime via NF-κB due to a mutation in *CARD11*. Stratifying patients by MSS revealed that patients with low synergy survived significantly longer than patients with high mutational synergy (Fig 4E, p = 0.02). By comparison, stratification of the same patient cohort by widely used gene expression-based cell-of-origin classifications and more recent mutational-based clustering did not reach significance for association with survival (Fig 4E). We thus demonstrated that using WES data to create

computational models of individual patients, and quantifying mutational synergy, enables prediction of OS in a cohort of patients with poor prognosis.

Discussion

Many types of cancer are clinically and genetically heterogeneous. Developments in sequencing technologies have led to the discovery of distinct subgroups of cancer based on mutational profiles. Often, as is the case in B-cell lymphoma, these disease subtypes can be differentiated by the signalling pathways that are most dysregulated (11-13). Despite this, knowledge of the signalling pathways and their interactions are not routinely used to classify disease or predict prognosis. Computational models of signalling networks can encode decades of knowledge (20), and many of these models were originally parameterised or validated in wet-lab experiments where genes are mutated (36). Here we demonstrated that by mapping genetic data onto computational models we could quantitatively predict the degree to which combinations of mutations combine deleteriously. We simulated mutations and chromosomal abnormalities by adjusting appropriate model parameters and thereby placed the genes effected in the context of molecular signalling networks. We also developed a novel metric: “mutational synergy score”, or MSS, to quantify how genomic events combine to contribute to disease severity. We demonstrated the ability of mutational synergy to identify established subgroups of lymphoma and myeloma that have the worst prognosis. This approach enabled us to create virtual patients using WES data and predict prognosis in patients by identifying the most deleterious combinations of mutations, even when clustering patients by their gene expression or mutational burden was not prognostically informative.

We used established multi-scale, agent-based models of B cells (18, 19), in which cellular signalling networks were encoded in continuous ODE models. Logical modelling is frequently used to simulate networks of this size and complexity, and indeed has been previously applied to B-cell lymphoma (37)**.** Here, through leveraging a continuous modelling framework, we were

able to identify many dose-dependent effects, including how chromosomal gain and amplification confer distinct prognoses. Given the size of the ODE systems used here (194 equations, and 563 parameters) a large amount of information is required to parameterise these ODE models. Here we performed no parameter fitting, as these parameters have been accumulated and validated across multiple cellular contexts and cell types (18, 20, 21, 36). We assume that parameters, other than those effected by mutations, remain consistent between healthy B-cells and cancerous B-cells. Indeed, many of these parameters have been shown to be consistent in multiple cell types including T-cells, dendritic cells and fibroblasts (20). Embedding Large ODE systems in agent-based models, that result in simulations of thousands of cells, presents a substantial computational challenge. Here, these challenges were overcome through high performance computing and by leveraging results that demonstrate that B-cell fates are reliably inherited (18), enabling simulation of a single branch down each lineage tree for each founder cell and interference of the fate of all remaining daughter cells. We also utilised the fact that individual mutations are present in multiple patients to reduce the need to simulate each individual mutation in each individual patient. We benefited from the performance of the Julia programming language (38), the DifferentialEquations package (39), and used Docker containers to ensure reproducibility.

Whilst we find computational modelling aligns with and exceeds clinically utilised disease classifications for these poor prognosis patients, further work is required to develop this approach into a clinical tool. The correlation between mutational synergy and overall survival is not strong enough alone to inform clinical prognostic decisions. This may result from some of the assumptions we needed to make, including assuming that all mutations have the same effect size, equivalent to one extra copy or loss of copy of each gene. We expect that identifying and quantifying the impact of each mutation on each gene would improve the correlation between mutational synergy and OS. Given that only WES data were used to inform the model,

and the assumptions made to map mutations to model parameters, the correlation between mutational synergy and outcome is perhaps surprising. Simulating the effect of targeted inhibitors in the patient-specific simulations presented here may provide insights into which patients will respond to these drugs, identify new treatment strategies, and shed light on treatment resistance.

Our analysis is limited to poor prognosis patients who succumb to the disease within approximately 3 years. Further work is required to simulate how treatment may perturb biological networks harbouring mutations across long time scales, by simulating the effects of R-CHOP chemotherapy, and polatuzumab- or etoposide-containing regimes. This work may predict and provide mechanistic insight into the development of relapsed DLBCL. Modelling long-term responses will require consideration of the causes of relapse. Previous modelling studies have addressed this challenge by simultaneously simulating populations of malignant cells, unmutated cells, and premalignant cells to predict the time taken for malignant cells to re-emerge is predicted (40). Clonal expansion dynamics and cellular plasticity, which are often associated with increased aggressiveness and treatment resistance, are likely to contribute to relapse over longer time scales than considered here (41). Expression of epithelial-mesenchymal transition transcription factors can drive indolent type lymphoma to take on an aggressive phenotype (42). Regulation of these gene expression programs has been the subject of multiple mathematical modelling approaches (43). Another model, which simulates the transition of cancer cells between stem and non-stem states over time, predicted that plasticity increased tumour growth rate and invasion (44). Creating models that can predict long-term outcomes and the emergence of treatment resistance from diagnostically-available data will require reconciling the timescales between the work we present here and models of treatment resistance/cellular plasticity. Focusing on the poor-prognosis patients in DH lymphoma enabled us to predict prognosis using computational models focused on molecular

signalling, while also focusing this study on the largest unmet need in this disease. Similarly, gain(1q) MM represents a substantial unmet need, and our findings that computational modelling can also predict the correlation between outcome and copies of chromosome 1q seen in clinical trials indicates that this approach will have utility in multiple malignancies.

Beyond DLBCL and MM, the modelling methodology and synergy calculation described here can be applied to other tumour types when mutational data and comprehensive, experimentally validated simulations are available. Many of the genes dysregulated in DLBCL and MM harbour common cancer-associated mutations. It is well-established that c-Myc deregulation is involved in the development of a wide variety of cancers (45), Bcl2 is dysregulated in many malignancies including breast cancer and gastric carcinoma (46, 47)**,** while NF-κB is implicated in numerous cancers (48)**.** A mathematical model, simulating 17 cancer types, estimated that the number of carcinogenic mutations (hits) varies from two to eight depending on the cancer type (49). With the increasing availability of cancer genome data and the evolution of computational methods to identify the mutational burden of cancer patients, the approach described here is rapidly becoming feasible in numerous cancers. Mutational synergy metrics are likely to be most impactful in mutationally heterogeneous cancers, such as breast cancer, where heterogeneity challenges early diagnosis, treatment selection and prognosis prediction (50).

The goal of cancer research, beyond understanding how the disease develops, is to find effective treatments for all patients. As knowledge of human cancers has developed it has become clear that, although patterns may emerge, every patient's cancer is different. We therefore need better tools, not only to diagnose cancer, but to predict survival and response to treatment. Modern sequencing techniques have provided ever more data, but we have yet to fully utilise it. Computational models with the power to translate patient-specific mutation data into personalised treatment plans will be at the forefront of patient care. The work presented

here is a step towards this, and in future we hope to develop it further and demonstrate its potential in a range of malignancies.

Methods

Model generation

Ordinary differential equation (ODE) model files were generated from individual reaction, parameter and rate law tables using CSV2Julia, a Python script developed in-house (see https://github.com/SiFTW/CSV2JuliaDiffEq). Jupyter Notebooks with Julia code used to generate and run all computational models and plot output are available at https://github.com/SiFTW/norrisEtAl. Ordinary differential equation models were solved using DifferentialEquations.jl (38, 39). Full description of the models reactions, parameters, and rate laws are defined in the GitHub repository.

Cell cycle model

Each cell was simulated until either a limit cycle was hit, or the maximum time allowed for the simulation ($1x10^6$ minutes). We defined the limit cycle as when the maximum value of cell mass is the same, within 3 decimal places, for 5 consecutive cycles (https://github.com/SiFTW/norrisEtAl/CellCycle_WTvsMycOE.ipynb). Simulations of cell populations were performed using distributions of parameters representing expression and degradation of modelled molecular species as described previously with 11% coefficient of variance sampled from a normal distribution around the published parameter value truncated at 0 (18) (https://github.com/SiFTW/norrisEtAl/CellCycleParamDist.ipynb).

Apoptosis model

The model was initially run to steady state. The final values for each molecular species were then used (unmutated) or modified (to reflect mutations) as input to a time course phase. Values for PARP and cleaved PARP were extracted and the ratio of PARP to cleaved PARP was calculated for each time point in the simulation. Time to death was defined as the point at which this ratio was >0.1 (https://github.com/SiFTW/norrisEtAl/Apoptosis_Bcl2.ipynb, https://github.com/SiFTW/norrisEtAl/ApoptosisDist.ipynb).

Multi-scale model

The multi-scale model was generated using module definition files (reactions, rate laws and parameters) for individual modules (NF-κB signalling, cell cycle, apoptosis, and differentiation) as well as a set of linking reactions. Any reactions or parameters appearing in individual modules that overlap with linking reactions were replaced by the corresponding information from the linking files. Cell populations were simulated according to a set of pre-defined conditions including condition name, parameter(s) to modify and amount to modify the parameter(s) by. All these conditions were kept constant for all simulations presented here except for individual parameters changed to individual mutations.

Callbacks were used to modify the values of molecular species during the simulation upon reaching a set of pre-defined conditions. In the multi-scale model this allows cells to divide or die, and in the case of division introduces new simulations. The two callbacks used in multi-scale modelling were maintained from previous work (cleaved PARP >2500 = apoptosis, CDH1 >0.1 = cell division) (18, 19).

The introduction of two daughter cells with each cell division resulted in an infeasible simulation size in highly proliferative simulations, and cell fates have been shown to be reliably inherited in B-cells such that both daughter cells will achieve similar fates at similar times (16). Therefore, at each cell division we introduced a single daughter representing both daughter cells, with a

variable to track the number of cells represented by each daughter, using the formula: [cells represented] = $2^{(generation-1)}$.

Files that define the multi-scale model along with Jupyter notebooks to run the simulations described here are available in the GitHub repository.

Modelling patients

We restricted our analysis to mutations that impact the core simulated network of the multi-scale model using a gene list of recurrent mutations that are known to impact well-defined nodes of the model (gene list: https://github.com/SiFTW/norrisEtAl/geneList.txt). Querying this gene list through the cBioPortal we converted mutations to parameter value changes based on literature regarding the effect of each mutation. For gain-of-function mutations we multiplied the appropriate parameter value by 1.5 and for loss-of-function we multiplied by 0.5. For copy number alterations (CNA) we found each gene on the respective chromosome section and multiplied by 1.5 for an amplification and 0.5 for heterozygous loss. If multiple mutations affected the same parameter, then the individual effects of those mutations were multiplied together. For mutations in genes not modelled directly we assigned an indirect effect. For example, the effect of mutations in p53 is modelled as overexpression of *BCL2* as it is known to inhibit Bcl2, and *CARD11* mutations are modelled as increasing NEMO-IKK activity (51).

Calculating Metrics

For each simulation the number of cells alive in the population was exported, and we then calculated a series of metrics (see https://github.com/SiFTW/norrisEtAl/calculateCellNumbersForAllConditions.ipynb):

PeakCN = maximum number of cells at any time point in the simulation of a patients' full genotype

PeakEIA = maximum number of cells at any time point calculated from the sum of all individual mutations minus WT.

PeakSyn = maximum value of peakCN minus peakEIA

NormalisedSyn = peakSyn/peakEIA

Acknowledgements

The authors would like to thank Ulf Klein, Jana Wolf, and Richard Norris for critical discussions and Eleanor Jayawant for providing critical input into the manuscript. Funding for SM: Leukaemia UK John Goldman Fellowship (2020/JGF/003) and UKRI Future Leaders Fellowship (MR/T041889/1). Funding for AP: MRC Research Grant (MR/V009095/1). Funding for CP: Blood Cancer UK (23004).

Author contributions

RN and SM performed computational analysis and modelling. RN, CP, AP and SM conceptualized the study, and EM, JJ and TC conceptualized the Figure 3. AP, CP and SM supervised the project. All authors contributed to the article, critically revised the manuscript, and approved the submitted version.

Disclosure and competing interests statement

The authors state they have no competing interests or disclosures.

Data Availability

The datasets and computer code produced in this study are available in the following databases:

Modelling computer scripts: GitHub (https://github.com/SiFTW/norrisEtAl).

438

439

440

441

442

References

1. Maura F, Bolli N, Angelopoulos N, Dawson KJ, Leongamornlert D, Martincorena I, et al. Genomic landscape and chronological reconstruction of driver events in multiple myeloma. Nat Commun. 2019;10(1):3835.
2. Mosquera Orgueira A, Ferreiro Ferro R, Díaz Arias J, Aliste Santos C, Antelo Rodríguez B, Bao Pérez L, et al. Detection of new drivers of frequent B-cell lymphoid neoplasms using an integrated analysis of whole genomes. PLoS One. 2021;16(5):e0248886.
3. Cheah C, Herbert K, O'rourke K, Kennedy GA, George A, Fedele P, et al. A multicentre retrospective comparison of central nervous system prophylaxis strategies among patients with high-risk diffuse large B-cell lymphoma. British journal of cancer. 2014;111(6):1072-9.
4. Hartert KT, Wenzl K, Krull JE, Manske M, Sarangi V, Asmann Y, et al. Targeting of inflammatory pathways with R2CHOP in high-risk DLBCL. Leukemia. 2021;35(2):522-33.
5. Walz S, Lorenzin F, Morton J, Wiese KE, von Eyss B, Herold S, et al. Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles. Nature. 2014;511(7510):483-7.
6. Xu-Monette ZY, Deng Q, Manyam GC, Tzankov A, Li L, Xia Y, et al. MYC mutation profiling and prognostic significance in de novo diffuse large B-cell lymphoma. Clinical cancer research: an official journal of the American Association for Cancer Research. 2016;22(14):3593.
7. Yip K, Reed J. Bcl-2 family proteins and cancer. Oncogene. 2008;27(50):6398-406.
8. Hanahan D, Weinberg RA. The hallmarks of cancer. cell. 2000;100(1):57-70.
9. Ci W, Polo JM, Cerchietti L, Shaknovich R, Wang L, Yang SN, et al. The BCL6 transcriptional program features repression of multiple oncogenes in primary B cells and is deregulated in DLBCL. Blood, The Journal of the American Society of Hematology. 2009;113(22):5536-48.
10. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. New England Journal of Medicine. 2002;346(25):1937-47.
11. Chapuy B, Stewart C, Dunford AJ, Kim J, Kamburov A, Redd RA, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nat Med. 2018;24(5):679-90.
12. Lacy SE, Barrans SL, Beer PA, Painter D, Smith AG, Roman E, et al. Targeted sequencing in DLBCL, molecular subtypes, and outcomes: a Haematological Malignancy Research Network report. Blood. 2020;135(20):1759-71.
13. Schmitz R, Wright GW, Huang DW, Johnson CA, Phelan JD, Wang JQ, et al. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. N Engl J Med. 2018;378(15):1396-407.
14. Ennishi D, Jiang A, Boyle M, Collinge B, Grande BM, Ben-Neriah S, et al. Double-hit gene expression signature defines a distinct subgroup of germinal center B-cell-like diffuse large B-cell lymphoma. Journal of Clinical Oncology. 2019;37(3):190.
15. Sehn LH, Salles G. Diffuse Large B-Cell Lymphoma. N Engl J Med. 2021;384(9):842-58.
16. Liu X, Jia S, Chu Y, Tian B, Gao Y, Zhang C, et al. Chromosome 1q21 gain is an adverse prognostic factor for newly diagnosed multiple myeloma patients treated with bortezomib-based regimens. Front Oncol. 2022;12:938550.
17. Abdallah N, Greipp P, Kapoor P, Gertz MA, Dispenzieri A, Baughn LB, et al. Clinical characteristics and treatment outcomes of newly diagnosed multiple myeloma with chromosome 1q abnormalities. Blood Adv. 2020;4(15):3509-19.
18. Mitchell S, Roy K, Zangle TA, Hoffmann A. Nongenetic origins of cell-to-cell variability in B lymphocyte proliferation. Proceedings of the National Academy of Sciences. 2018;115(12):E2888-E97.
19. Roy K, Mitchell S, Liu Y, Ohta S, Lin Y-s, Metzig MO, et al. A regulatory circuit controlling the dynamics of NFκB cRel transitions B cells from proliferation to plasma cell differentiation. Immunity. 2019;50(3):616-28. e6.

491 20. Mitchell S, Tsui R, Tan ZC, Pack A, Hoffmann A. The NF-κB multidimer system model: A
492 knowledge base to explore diverse biological contexts. Science Signaling. 2023;16(776):eabo2838.
493 21. Shokhirev MN, Almaden J, Davis-Turak J, Birnbaum HA, Russell TM, Vargas JA, et al. A multi-scale
494 approach reveals that NF-κB cRel enforces a B-cell decision to divide. Mol Syst Biol. 2015;11(1):783.
495 22. Youle RJ, Strasser A. The BCL-2 protein family: opposing activities that mediate cell death. Nat
496 Rev Mol Cell Biol. 2008;9(1):47-59.
497 23. Heinzel S, Binh Giang T, Kan A, Marchingo JM, Lye BK, Corcoran LM, et al. A Myc-dependent
498 division timer complements a cell-death timer to regulate T cell and B cell responses. Nat Immunol.
499 2017;18(1):96-103.
500 24. Loriaux PM, Tesler G, Hoffmann A. Characterizing the relationship between steady state and
501 response using analytical expressions for the steady states of mass action models. PLoS computational
502 biology. 2013;9(2):e1002901.
503 25. Dowling MR, Kan A, Heinzel S, Zhou JH, Marchingo JM, Wellard CJ, et al. Stretched cell cycle
504 model for proliferating lymphocytes. Proc Natl Acad Sci U S A. 2014;111(17):6377-82.
505 26. Cai C, Hu X, Dai P, Zhang T, Jiang M, Wang L, et al. c-Myc regulates neural stem cell quiescence
506 and activation by coordinating the cell cycle and mitochondrial remodeling. Signal Transduct Target
507 Ther. 2021;6(1):306.
508 27. Cheng Y, Luo H, Izzo F, Pickering BF, Nguyen D, Myers R, et al. m(6)A RNA Methylation Maintains
509 Hematopoietic Stem Cell Identity and Symmetric Commitment. Cell Rep. 2019;28(7):1703-16.e6.
510 28. Duffy KR, Hodgkin PD. Intracellular competition for fates in the immune system. Trends Cell Biol.
511 2012;22(9):457-64.
512 29. Meriranta L, Pasanen A, Alkodsi A, Haukka J, Karjalainen-Lindsberg M-L, Leppä S. Molecular
513 background delineates outcome of double protein expressor diffuse large B-cell lymphoma. Blood
514 advances. 2020;4(15):3742-53.
515 30. El-Sharkawi D, Sud A, Prodger C, Khwaja J, Shotton R, Hanley B, et al. A retrospective study of
516 MYC rearranged diffuse large B-cell lymphoma in the context of the new WHO and ICC classifications.
517 Blood Cancer J. 2023;13(1):54.
518 31. Quesada AE, Medeiros LJ, Desai PA, Lin P, Westin JR, Hawsawi HM, et al. Increased MYC copy
519 number is an independent prognostic factor in patients with diffuse large B-cell lymphoma. Mod Pathol.
520 2017;30(12):1688-97.
521 32. Schmidt TM, Fonseca R, Usmani SZ. Chromosome 1q21 abnormalities in multiple myeloma.
522 Blood Cancer J. 2021;11(4):83.
523 33. Hanamura I. Gain/Amplification of Chromosome Arm 1q21 in Multiple Myeloma. Cancers
524 (Basel). 2021;13(2).
525 34. Sarkozy C, Traverse-Glehen A, Coiffier B. Double-hit and double-protein-expression lymphomas:
526 aggressive and refractory lymphomas. The Lancet Oncology. 2015;16(15):e555-e67.
527 35. Akyurek N, Uner A, Benekli M, Barista I. Prognostic significance of MYC, BCL2, and BCL6
528 rearrangements in patients with diffuse large B-cell lymphoma treated with cyclophosphamide,
529 doxorubicin, vincristine, and prednisone plus rituximab. Cancer. 2012;118(17):4173-83.
530 36. Hoffmann A, Levchenko A, Scott ML, Baltimore D. The IkappaB-NF-kappaB signaling module:
531 temporal control and selective gene activation. Science. 2002;298(5596):1241-5.
532 37. Thobe K, Konrath F, Chapuy B, Wolf J. Patient-Specific Modeling of Diffuse Large B-Cell
533 Lymphoma. Biomedicines. 2021;9(11).
534 38. Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: A fresh approach to numerical computing.
535 SIAM review. 2017;59(1):65-98.
536 39. Rackauckas C, Nie Q. Differentialequations. jl–a performant and feature-rich ecosystem for
537 solving differential equations in julia. Journal of open research software. 2017;5(1).

538 40. Abubakar SD, Takaki M, Haeno H. Computational modeling of locoregional recurrence with
539 spatial structure identifies tissue-specific carcinogenic profiles. Front Oncol. 2023;13:1116210.
540 41. Qin S, Jiang J, Lu Y, Nice EC, Huang C, Zhang J, et al. Emerging role of tumor cell plasticity in
541 modifying therapeutic response. Signal Transduct Target Ther. 2020;5(1):228.
542 42. Chen SC, Liao TT, Yang MH. Emerging roles of epithelial-mesenchymal transition in
543 hematological malignancies. J Biomed Sci. 2018;25(1):37.
544 43. Tripathi S, Xing J, Levine H, Jolly MK. Mathematical Modeling of Plasticity and Heterogeneity in
545 EMT. Methods Mol Biol. 2021;2179:385-413.
546 44. Poleszczuk J, Enderling H. Cancer Stem Cell Plasticity as Tumor Growth Promoter and Catalyst of
547 Population Collapse. Stem Cells Int. 2016;2016:3923527.
548 45. Kalkat M, De Melo J, Hickman KA, Lourenco C, Redel C, Resetca D, et al. MYC Deregulation in
549 Primary Human Cancers. Genes (Basel). 2017;8(6).
550 46. Joensuu H, Pylkkänen L, Toikkanen S. Bcl-2 protein expression and long-term survival in breast
551 cancer. Am J Pathol. 1994;145(5):1191-8.
552 47. Inada T, Kikuyama S, Ichikawa A, Igarashi S, Ogata Y. Bcl-2 expression as a prognostic factor of
553 survival of gastric carcinoma. Anticancer Res. 1998;18(3b):2003-10.
554 48. Luo JL, Kamata H, Karin M. IKK/NF-kappaB signaling: balancing life and death--a new approach
555 to cancer therapy. J Clin Invest. 2005;115(10):2625-32.
556 49. Anandakrishnan R, Varghese RT, Kinney NA, Garner HR. Estimating the number of genetic
557 mutations (hits) required for carcinogenesis based on the distribution of somatic mutations. PLoS
558 Comput Biol. 2019;15(3):e1006881.
559 50. Guo L, Kong D, Liu J, Zhan L, Luo L, Zheng W, et al. Breast cancer heterogeneity and its
560 implication in personalized precision therapy. Exp Hematol Oncol. 2023;12(1):3.
561 51. Hemann MT, Lowe SW. The p53-Bcl-2 connection. Cell Death Differ. 2006;13(8):1256-9.

562

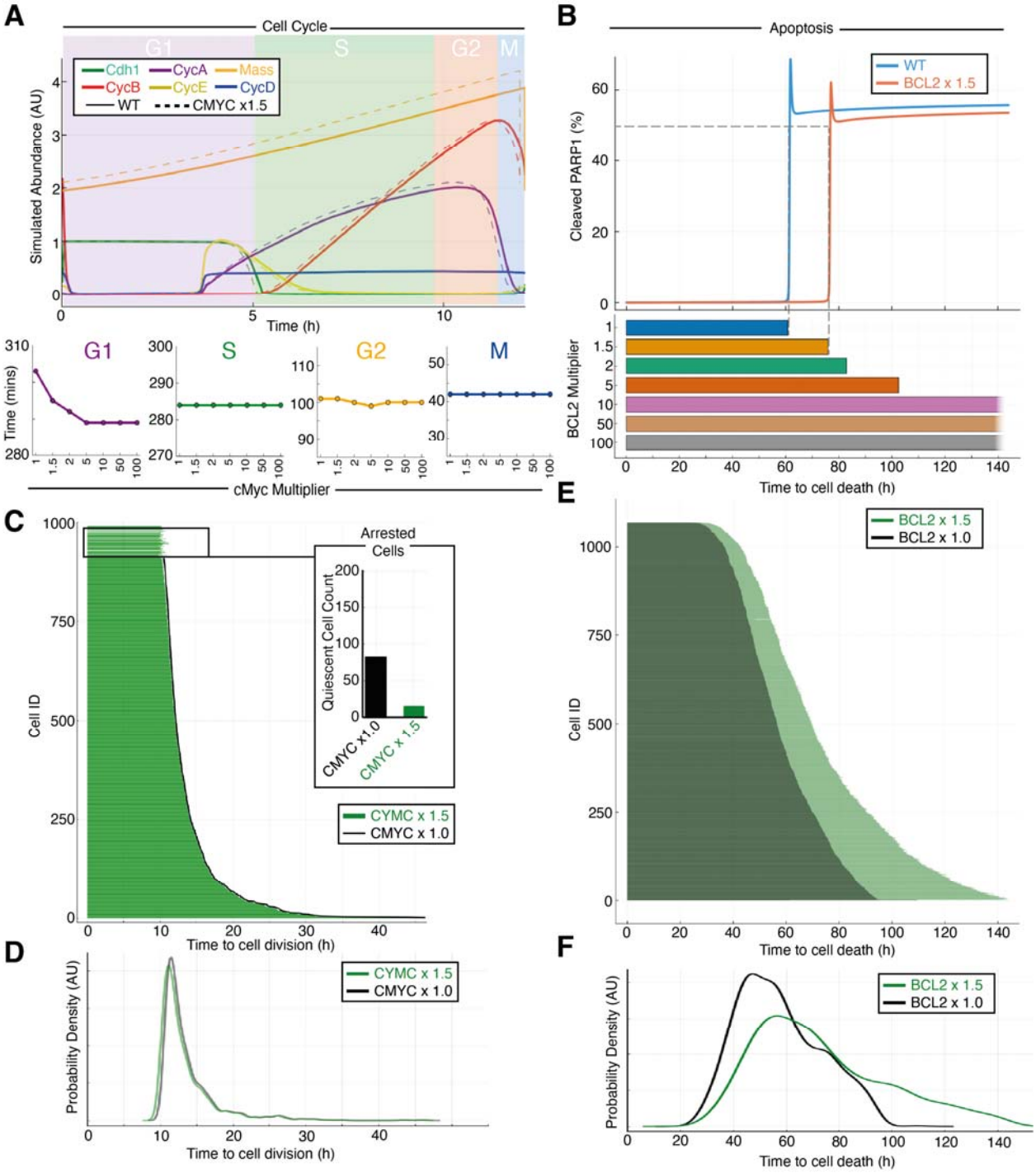

Figure 1. Computational modelling of the cell cycle and apoptosis reveal limited impact of archetypal double hit mutations on their respective molecular networks.

A. (top) Output of the cell cycle model run to a limit cycle for a WT cell and a cell in which *MYC* is overexpressed (x1.5). Lines indicate the values of molecular species (used to calculate phase transition times) over time from the first growth phase (G1) to cell division (M). Cyclin is abbreviated to “Cyc”.

A. (bottom) Cell cycle phase timing for cells with varying *MYC* expression.

B. (top) The percentage of cleaved PARP over time for two simulations (WT and *BCL2* overexpressed) using the apoptosis model. B (bottom). Time to death for cells with varying *BCL2* expression.

C, D. Comparison of total cell cycle times from limit cycles of simulations of heterogeneous unmutated (c-MYC x1.0) cells and cells with a *MYC* mutation (c-MYC x1.5). C. Total cell cycle times in c-MYC x1.0 simulations are indicated with a black line, while c-MYC x1.5 is indicated with green bars. Each unmutated cell is plotted alongside the corresponding *MYC* mutant (by cell ID). The only difference between matched cells is *MYC* expression level but each cell pair has distributed parameters compared to every other pair. The inset bar plot indicates the number of cells (in c-MYC x1.0 & x1.5 simulations) in which cell cycle arrest is predicted. D. Cell cycle times for the same simulations represented as panel C but displayed density plots. c-MYC x1.0 is shown in black and c-MYC x1.5 in green.

E, F. Comparisons of time to death for simulations of heterogeneous populations of matched cells differing only in expression level of *BCL2*. Simulations are shown as both a horizontal bar plot (E) with BCL2 x1.0 shown in black and BCL2 x1.5 shown in green, and a density plot (F) with the same colour scheme.

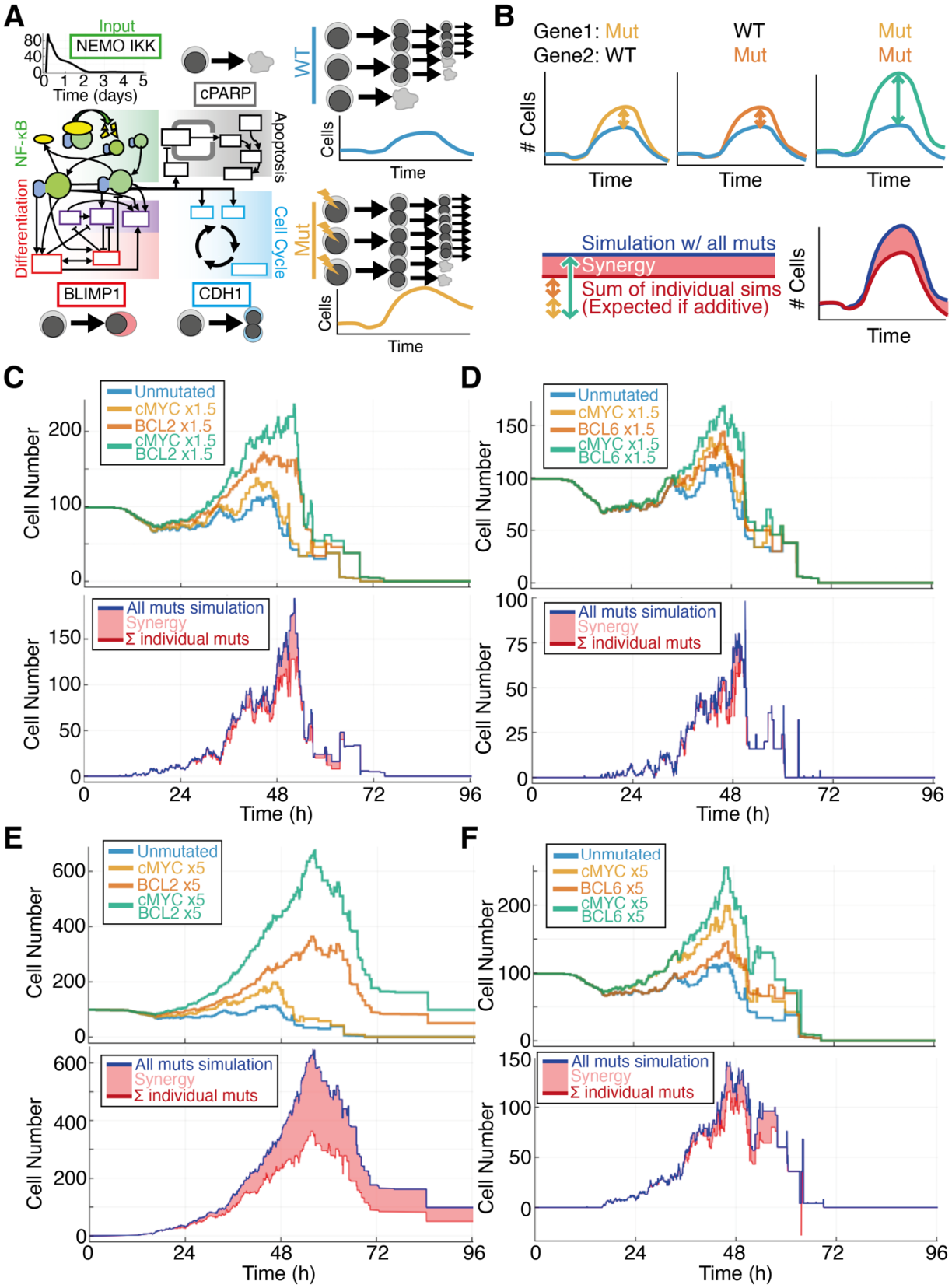
A
Input
NEMO IKK
80
40
0
0 1 2 3 4 5
Time (days)
cPARP
NF-κB
Apoptosis
Differentiation
Cell Cycle
BLIMP1
CDH1
WT
Mut
Cells
Time
B
Gene1: Mut
Gene2: WT
WT
Mut
Mut
Mut
Cells
Time
Simulation w/ all muts
Synergy
Sum of individual sims
(Expected if additive)
C
Unmutated
cMYC x1.5
BCL2 x1.5
cMYC x1.5
BCL2 x1.5
Cell Number
All muts simulation
Synergy
Σ individual muts
Time (h)
D
Unmutated
cMYC x1.5
BCL6 x1.5
cMYC x1.5
BCL6 x1.5
E
Unmutated
cMYC x5
BCL2 x5
cMYC x5
BCL2 x5
F
Unmutated
cMYC x5
BCL6 x5
cMYC x5
BCL6 x5

Figure 2. Multi-scale modelling of heterogeneous cell populations reveals that mutations combine with oncogenic synergy.

A) Schematics of the multi-scale model, including the signalling networks within the model and the cell fates they control. The full reaction scheme of the model is defined in the Github repository associated with this study: https://github.com/SiFTW/norrisEtAl. Simulation input, NEMO and IKK activity curves, are indicated. A schematic tree structures of cell division and death in simulations of cell populations with and without mutations is provided (right).

B) Explanation of the method for calculating a comparison of individual mutations and combinations of mutations (mutational synergy). Mutational synergy is defined as the difference in cell number between the sum of cell numbers from all simulations containing individual mutations, and the cells added in a simulation that combines all mutations. This difference defines the red region in this schematic and future plots.

C-F) Cell population size (cell count) over time for wild type/unmutated (blue), individual mutations (orange and red) and double hits (green). Each simulation started with 100 heterogeneous B-cells stimulated by the indicated IKK input curve (panel A). C & E) Cell numbers and synergistic regions for *MYC* and *BCL2* overexpression x1.5 (C) and x5 (E). D & F) *MYC* and *BCL6* overexpression x1.5 (D) and x5 (F).

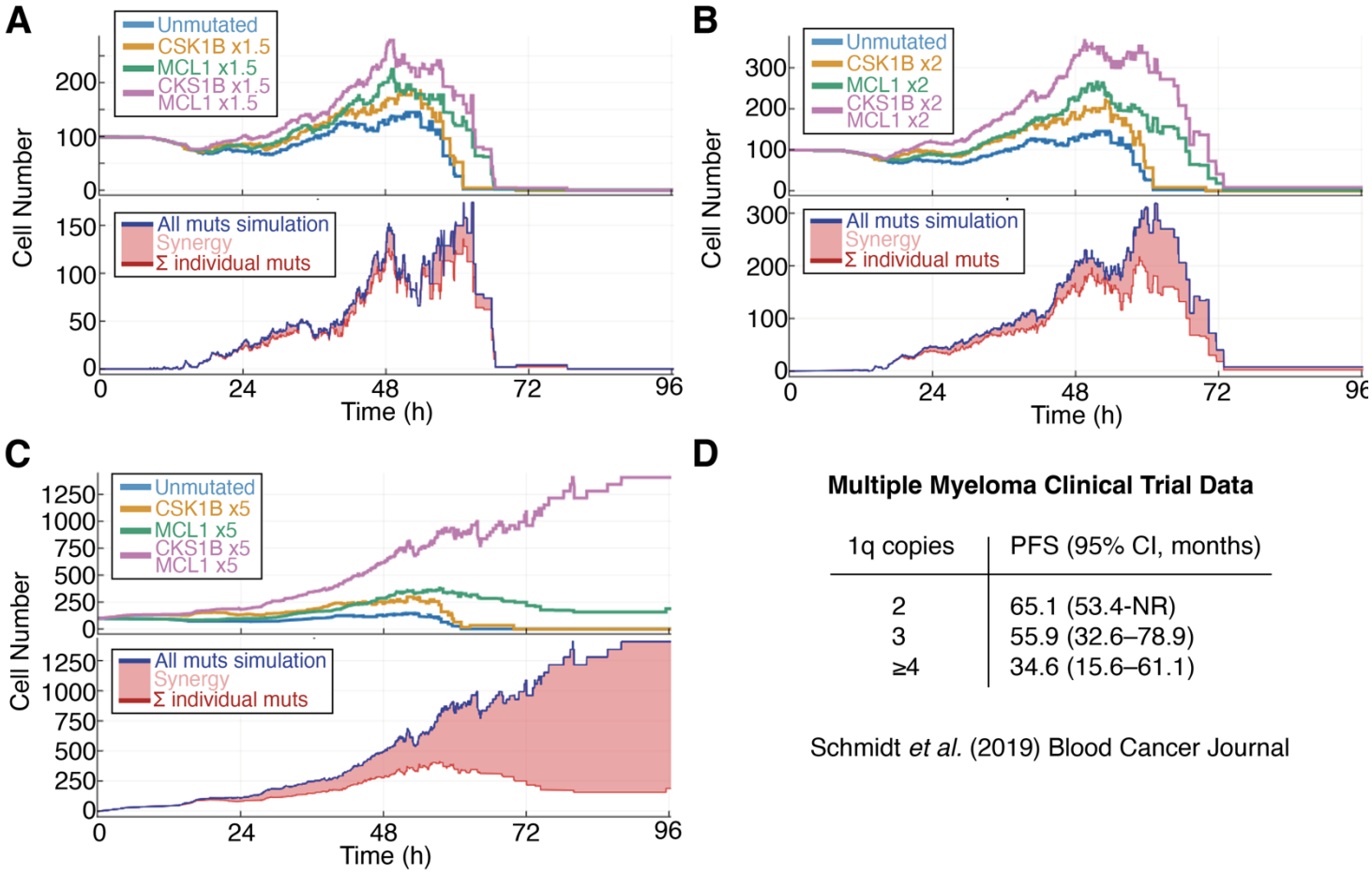

Figure 3. Mutations associated with aggressive multiple myeloma are also predicted to create mutational synergy.

A-C) Cell population size (cell count) over time for wild type/unmutated (blue), individual mutations (orange and green) and double hits (pink). *CKS1B* and *MCL1* overexpression x1.5 (A), x2 (B) and x5 (C).

D) Data from Schmidt et al., (32) showing the average progression free survival (PFS) for patients in a Multiple Myeloma cohort with varying chromosome 1q copy number.

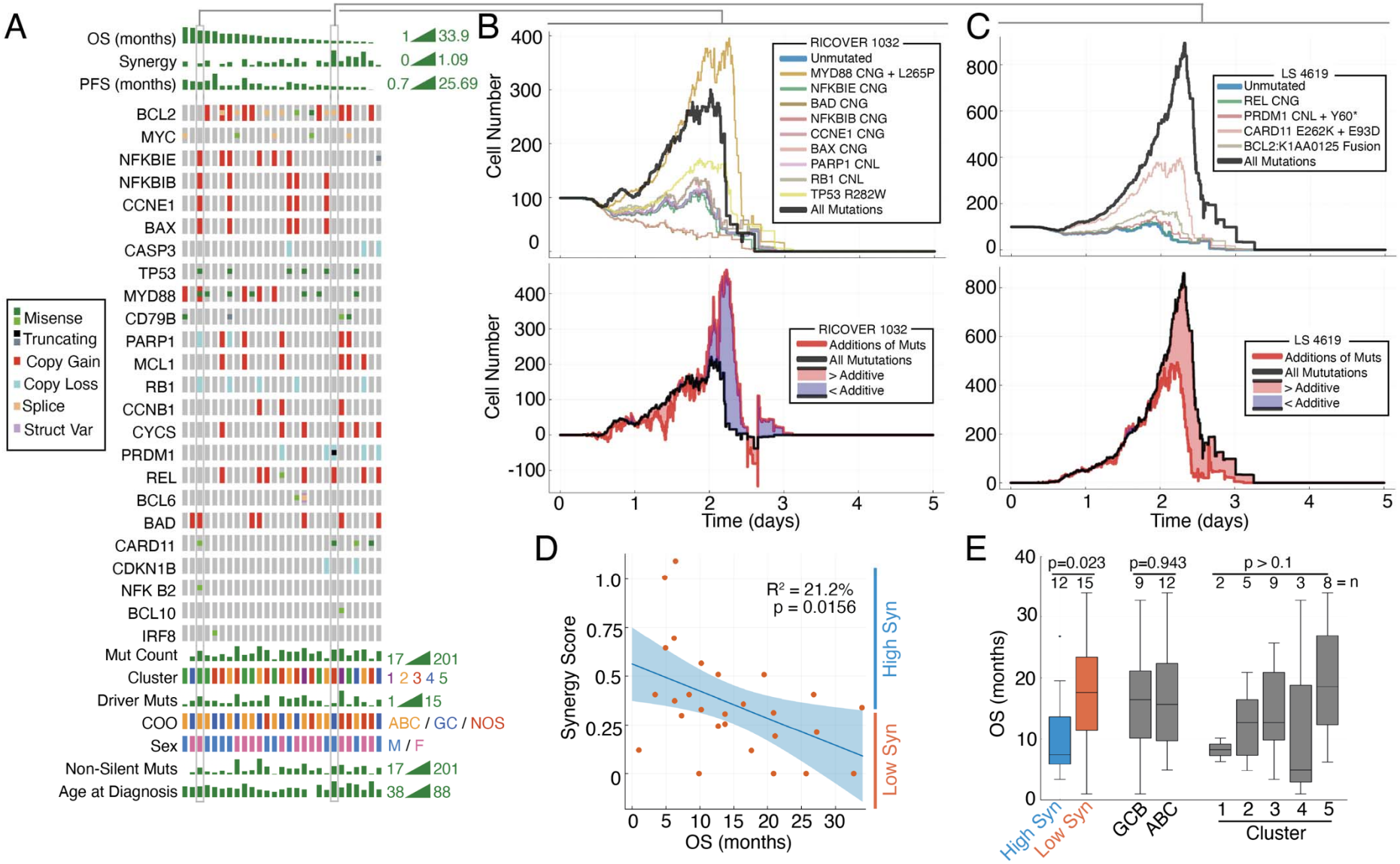
A
OS (months)
Synergy
PFS (months)
1 33.9
0 1.09
0.7 25.69
BCL2
MYC
NFKBIE
NFKBIB
CCNE1
BAX
CASP3
TP53
MYD88
CD79B
PARP1
MCL1
RB1
CCNB1
CYCS
PRDM1
REL
BCL6
BAD
CARD11
CDKN1B
NFK B2
BCL10
IRF8
Mut Count 17 201
Cluster 1 2 3 4 5
Driver Muts 1 15
COO ABC / GC / NOS
Sex M / F
Non-Silent Muts 17 201
Age at Diagnosis 38 88
Misense
Truncating
Copy Gain
Copy Loss
Splice
Struct Var
B
Cell Number
RICOVER 1032
Unmutated
MYD88 CNG + L265P
NFKBIE CNG
BAD CNG
NFKBIB CNG
CCNE1 CNG
BAX CNG
PARP1 CNL
RB1 CNL
TP53 R282W
All Mutations
Additions of Muts
All Mututations
> Additive
< Additive
Time (days)
C
LS 4619
Unmutated
REL CNG
PRDM1 CNL + Y60*
CARD11 E262K + E93D
BCL2:K1AA0125 Fusion
All Mutations
Additions of Muts
All Mututations
> Additive
< Additive
Time (days)
D
Synergy Score
R² = 21.2%
p = 0.0156
High Syn
Low Syn
OS (months)
E
OS (months)
p=0.023
p=0.943
p > 0.1
12 15
9 12
2 5 9 3 8 = n
High Syn
Low Syn
GCB
ABC
1 2 3 4 5
Cluster

Figure 4. Mutational synergy is correlated with overall survival in patients with the worst outcomes.

A) OncoPrint output showing mutations and clinical data for deceased patients from Chapuy et al., (11) with overall survival <36 months. All data is from Chapuy et al., (11) except mutational synergy which is calculated here.

B, C) Plots of cell number vs time for simulations of individual mutations and all mutations combined for patients with low (B) and high (C) synergy.

D) Plot of normalised synergy score vs overall survival for deceased patients with OS <36 months. The blue line shows best fit for synergy vs OS correlation; blue shadowing shows 95% confidence interval; orange and blue bars (right) delineate groups of patients with low and high synergy.

E) Box plots showing mean survival (months) for groups of patients with high (blue) and low (orange) synergy. Adjacent box plots show how stratifying patients by cell of origin and mutational clusters compares to the normalised synergy metric when plotted against overall survival (OS). P> 0.1 for all pairwise comparisons between mutational clusters for these patients from Chapuy et al. (11).