Patient-specific computational models predict prognosis in B cell lymphoma by quantifying pro-proliferative and anti-apoptotic signatures from genetic sequencing data.

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: Personalised lymphoma models predict prognosis.

Abstract

Genetic heterogeneity and co-occurring driver mutations impact clinical outcomes in blood cancers. Grouping tumours into clusters based on genetic alterations is prognostically informative. However, predicting the emergent effect of co-occurring mutations that impact multiple complex and interacting signalling networks remains challenging. Here, we used mathematical models to predict the impact of co-occurring mutations on cellular signalling and cell fates in diffuse large B cell lymphoma (DLBCL) and multiple myeloma (MM). Simulations predicted adverse impact on clinical prognosis when combinations of mutations induced both pro-proliferative and anti-apoptotic signalling. So, we established a pipeline to integrate patient-specific mutational profiles into personalised lymphoma models. Using this approach, we identified a subgroup (19%) of patients characterised by simultaneous upregulation of anti-apoptotic and pro-proliferative (AAPP) signalling. AAPP patients have dismal prognosis and can be identified within all current genomic and cell-of-origin classifications. Combining personalised

molecular simulations with mutational clustering enabled stratification of patients into clinically informative prognostic categories: good (80% progression-free survival at 120 months), intermediate (median progression-free survival of 93 months), and poor (AAPP, median progression-free survival of 26 months). This study shows that personalised computational models enable identification of novel high-risk patient subgroups, providing a valuable tool for future risk-stratified clinical trials.

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

Introduction

Mutational heterogeneity in haematological malignancies represents a major barrier to understanding the molecular mechanisms responsible for treatment response, 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 (1, 2). This heterogeneity contributes to variation in response to current treatments.

The most aggressive haematological malignancies frequently contain genetic aberrations affecting multiple genes, either through co-occurring mutations, e.g. double hit (DH) DLBCL or changes in the copy number of chromosomal regions containing multiple genes, e.g. gain1q multiple myeloma. DH DLBCL, featuring overexpression of *MYC* and *BCL2* (or *BCL6*), is among the most aggressive lymphoid malignancies, with very poor patient outcomes (3, 4). However, DH lymphoma represents fewer than 10% of DLBCL cases, while many more (30-40%) DLBCL patients relapse following frontline treatment (5). New approaches are needed to prospectively

45 identify poor prognosis patients with the aim of developing more effective treatment strategies
46 for this group.

47 Gene expression profiling split DLBCL into subgroups based on their putative cell of origin
48 (germinal centre – GC, or activated B cell – ABC) (6). Subsequent studies leveraged WES to
49 identify 5 or more prognostically informative patient clusters (7-9). These genetic groups have
50 been arbitrarily named cluster 1 to 5 (C1-C5) or with names referencing the most mutated
51 signalling pathways (MCD = MYD88 + CD79B), but are broadly consistent across multiple
52 studies (e.g. C5 aligns with MCD) (7-9). Clustering patients in this way largely ignores the
53 complexity of the molecular signalling networks that may be impacted by their individual genetic
54 landscapes. For example, a patient with most mutations converging on NF-κB is likely to be
55 assigned to C5 (7). However, this assignment ignores the presence or absence of functionally
56 significant co-occurring mutations in other pathways that may impact prognosis.

57 Computational models of molecular signalling in normal B cells have been used to predict cell
58 proliferation and survival; predictions that have been validated by in vitro laboratory experiments
59 with single cell resolution (10-13). Furthermore, incorporating mutations, and the impact of
60 mutations on protein abundance/activity, into these models predicts cellular responses in
61 experimental assays (10, 11, 14, 15). However, it is not known whether these models can
62 predict outcomes at the individual patient scale, nor whether mutational data alone is sufficient
63 to enable *in silico* simulations to make clinically relevant predictions in blood cancers. We
64 hypothesise that contextualising sequencing data within patient-specific, virtual signalling
65 networks may more precisely delineate the consequent cell fate decisions that impact
66 prognosis.

67 In this study we used mechanistic computational models to simulate how mutations combine in
68 B cell malignancies, and determined whether these approaches could predict patient outcome.

69 We developed a pipeline to create individual patient simulations using WES data, and tested
70 whether these personalised models could generate clinically informative prognostic information.

71 Materials and Methods

72 Detailed computational methods, and a lay summary of the methods to generate all
73 computational figures, are provided in the Supplementary Material. Methods are summarised
74 below. Code (as Jupyter notebooks) used to generate and run all computational models,
75 including descriptions of the reactions, parameters, and rate laws of each model, and code to
76 plot output are available in the Github repository (https://github.com/SiFTW/norrisEtAl).

77 Model generation

78 We employed established, computational models of healthy B cells, which enable simulation of
79 proliferation, apoptosis and terminal B cell differentiation in a heterogeneous B cell population
80 (10, 11, 13). The models consisted of a series of differential equations, representing the rate of
81 change of biomolecules (mRNAs, proteins, or protein complexes) over time. Models were
82 converted to Julia to be solved using DifferentialEquations.jl (16, 17).

83 Cell cycle and apoptosis modelling

84 The molecular networks representing apoptosis and the cell cycle were isolated from the
85 comprehensive B cell model (10, 11, 18). Cell-to-cell variability was simulated by distributing
86 model parameters as described previously (10). Over-expression of cMyc and Bcl2 was
87 simulated by increasing the parameter controlling the transcription rate of the respective mRNA.
88 An equilibrium phase was simulated prior to all time course simulations. For the cell cycle
89 model, the equilibrium phase ended when the same cell mass was reached prior to each cell
90 division for 5 cycle (within 3 decimal places). The final state of the equilibrium phase was used

91 as the initial condition for the time course phase. Cell death was defined as the first time point at
92 which over 10% of PARP was cleaved.

93 Multi-scale model

94 The NF-κB component of the established B cell model was updated to include reactions from a
95 more recent and comprehensive model of NF-κB (12). Apoptosis and cell division were
96 triggered as published previously (cleaved PARP >2500 triggers apoptosis, CDH1 >0.1 triggers
97 cell division) (10, 11). Previous studies introduced two separately simulated daughter cells with
98 each cell division (10, 11, 18), however here this resulted in an unfeasible simulation size in
99 highly proliferative simulations. Cell fates have been shown to be reliably inherited in B cells
100 such that daughter cells will achieve similar fates at similar times (16). Therefore, at each cell
101 division we replaced the mother cell with a single simulation representing $2^{(generation-1)}$ daughter
102 cells (2 cells following first division, 4 cells following second division etc.).

103 Modelling patients

104 Mutation profiles identified from WES and clinical data were downloaded through the cBioPortal
105 (7). We restricted the analysis to genetic changes impacting genes that could be mapped to
106 modelled parameters (https://github.com/SiFTW/norrisEtAl/blob/main/geneList.txt) and
107 leveraged OncoKB to simulate the impact of mutations annotated as “likely oncogenic” (31). We
108 created 113 patient-specific models (data for 135 patients available, 24 patients did not harbour
109 any mutations that mapped to model parameters). We assumed half of the normal expression
110 rate of each mRNA could be attributed to one of the two copies of the gene. Therefore,
111 parameters halved for copy number loss and loss-of-function mutations and increased by half
112 (1.5-fold change) for copy number gain and gain-of-function mutations. Copy number alterations
113 of chromosomal regions were modelled by identifying each gene that could be mapped to model

114 parameters within the chromosomal region. Multiple mutations affecting the same parameter
115 were combined multiplicatively, for example a patient with a MYD88 mutation and a CD79B
116 mutation (both increasing NEMO:IKK activity by 1.5 fold individually) was simulated as 2.25-fold
117 increased NEMO:IKK activity. The effect of mutations in genes not modelled directly was
118 assigned to the closest modelled molecular species: e.g. *MCL1* is modelled as overexpression
119 of *BCL2*, and *CARD11* mutations are modelled as increasing NEMO-IKK activity. As
120 biomolecules such as BCL2/cMYC mRNA represented the sum effect of multiple mRNAs, e.g.
121 MCL1+BCL2, these were renamed anti-apoptotic/pro-proliferative (AAPP) factors respectively,
122 to avoid inaccurate comparisons with experimental measurements of these transcripts. Patient
123 simulations were performed for 12 simulated hours, on a single cell with all parameters
124 consistent across all patients other than those impacted by mutations. The concentration of pro-
125 /anti-apoptotic factors at 6h was used for patient stratification, with patients stratified as above
126 or below the mean concentration at that time point.

127 ### Downstream analysis

128 All model outputs were analysed, and plots generated using the Julia programming language.
129 Scripts relating to each figure. can be found in the corresponding folder available in the GitHub
130 repository (https://github.com/SiFTW/norrisEtAl/).

131

132 ## Results

133 ### Simulating individual mutations recapitulates experimental measurements.

134 To determine whether established computational models of B cell molecular signalling could
135 make predictions of patient outcomes we first focused on DH DLBCL. The apoptotic regulatory

136 network was isolated from an established B cell model, and we simulated the impact of *BCL2*
137 overexpression on apoptotic signalling (11, 18)(Fig. 1A). A gene dose-dependent delay of
138 apoptosis measured by poly ADP-ribose polymerase (PARP) cleavage up to the equivalent of 5
139 extra copies of *BCL2* was seen in these simulations, beyond which the cells survive the full 140
140 hours of the simulation (Fig 1B). We repeated these simulations in a population of 1000
141 molecularly heterogeneous individual B cells (Fig. 1C and S1A). In these simulations cell death
142 times retained a log-normal distribution, but that 1.5-fold *BCL2* over-expression increased the
143 mean survival time in individual B cells by 1.3-fold (Fig. 1C). This is consistent with the delayed
144 cell death seen in B cells from Bcl-2-overexpressing mice (19).

145 *cMYC* is a proto-oncogene that plays a role in tumour cell growth and has previously been
146 modelled as acting on B cell progression within the cell cycle (13, 20). Simulating *cMYC* in this
147 way has been found to recapitulate both single-cell and bulk experimental measurements (10,
148 11). We isolated the cell cycle regulatory network from an established model of B cells and
149 simulated the impact of overexpression of *cMYC* (11, 18) (Fig. 1D). These simulations predicted
150 that increasing *cMYC* expression by the equivalent of 1 extra copy (1.5-fold increase) shortens
151 the cell cycle by just 8 minutes in a 12-hour cycle; an effect size much smaller than expected
152 (Fig. 1E). The sensitivity analysis also found that *cMYC* expression had a smaller effect on cell
153 cycle time than other genes in the pathway (Fig. S2). Importantly, this result recapitulated
154 experimental tracking of division times in B cells transfected with *cMYC* and cellular proliferation
155 measurement in lymphocytes from Eμ-Myc mice (21). In simulations of increasing levels of
156 *cMYC* expression, the small decrease in total cell cycle time was primarily due to a shortening
157 of G1 phase, with a gene dose-dependent effect up to the equivalent of 5-fold higher
158 expression, after which no further effect was seen (Fig. 1E). Upregulating *cMYC (x1.5)* in a
159 heterogeneous population of 1000 individual cells confirmed these findings and resulted in

160 simulated cell cycle timings that recapitulated the distribution of cell cycle duration measured by
161 time-lapse microscopy in murine B lymphocytes (22) (Fig. 1F and Fig. S1B). In these
162 simulations, increased *cMYC* expression was found to have a minor effect in all cells,
163 decreasing total average cell cycle time by just 11.5 minutes out of a 12-hour cycle (Fig 1F).
164 The wild-type simulation identified a population of cells (8%) in cell cycle arrest. Concentrations
165 of the cell cycle inhibitor p27, and complexes containing p27, were upregulated in these
166 arrested cells (Fig. S3), which recapitulates experimental measurements of cell cycle arrest
167 (24). Interestingly, in simulations of 1.5-fold *cMYC* overexpression the cell cycle arrest was
168 reversed in over 80% of these cells (Figure 1G). The number of cells rescued from arrest by
169 *cMYC* overexpression was proportional to the level of *cMYC* expression (Fig. S4) and these
170 rescued cells became the most rapidly proliferating cells within the cell population (Fig. 1G and
171 S1B). In keeping with these findings, *cMYC* is amongst the most differentially expressed genes
172 in quiescence and reverses quiescence in haematopoietic stem cells (23, 24). Recent data has
173 demonstrated that increasing the abundance of cMyc can enable senescent cancer cells to
174 resume division (25). Taken together this data shows that computational modelling of the impact
175 of *cMYC* and *BCL2* mutations, on the cell cycle and apoptosis respectively, accurately
176 recapitulates multiple experimental measurements.

177

178 Multi-scale modelling predicts that mutations that converge on apoptosis and the cell
179 cycle confer poor prognosis in blood cancer patients.

180 To establish whether computational simulations can accurately predict patient outcomes, we
181 initially simulated the well-characterised subgroup of lymphoma patients with mutations affecting
182 *cMYC* and *BCL2* (DH lymphoma). An agent-based multiscale model, previously used to
183 simulate immune responses, was used here (10, 11, 13). In this framework NF-κB-activation

184 through NEMO-IKK provided the input, resulting in the induction of NF-κB target genes (e.g.
185 CYCD, cMYC, IRF4, BCL2). The impact of these target genes on the molecular networks
186 controlling cell division and death could trigger the cell cycle to complete mitosis (resulting in
187 daughter cells being added) or the cell to undergo apoptosis (resulting in the cell being
188 removed, Fig. 2A and S5, see methods). Given that normally two copies of each gene are
189 present we assumed half (0.5-fold) the gene expression could be attributed to each copy.
190 Simulations with upregulated (equivalent to 1 extra copy, 1.5-fold increased expression) *cMYC*
191 and *BCL2*, resulted in cell number increases due to increased cell division and survival
192 respectively. When both genes were upregulated (1.5-fold expression) cell counts increased in
193 an additive manner (Fig. 2B, left). However, a 4-6 fold increase in expression of *BCL2* and
194 c*MYC* is commonly observed in these genes when they are mutated in lymphoma patient
195 samples (26). Under these conditions (5-fold increased expression), the simulations resulted in
196 an exponential increase in cell numbers (Fig. 2B, right). The classification of DH lymphoma
197 includes mutations in *cMYC* and *BCL2* or *BCL6*. Performing the same simulation on the *cMYC*
198 + *BCL6* DH lymphoma resulted in a substantially smaller effect compared to the *cMYC* + *BCL2*
199 DH (Fig. 2C). Using cell numbers as a proxy for disease prognosis would result in predictions
200 that *cMYC* + *BCL2* DH lymphoma would have worse prognosis than *cMYC* + *BCL6* DH
201 lymphoma (Fig. 2B and C). A multicentre retrospective data collection study, published while
202 this study was being conducted, confirmed our model prediction by showing that *cMYC* + *BCL6*
203 DH lymphoma does not represent high risk lymphoma, while *cMYC* + *BCL2* DH lymphoma has
204 particularly poor prognosis (Fig. 2D).

205

206 To determine whether computational predictions could recapitulate disease prognosis in other
207 high-risk haematological malignancies we simulated gain 1q multiple myeloma (Fig. 2E). Gain

208 1q MM is a commonly occurring cytogenetic event in MM associated with therapeutic failure and
209 inferior prognosis (27). Genes encoding the Bcl2-family protein Mcl1 and cell cycle regulator
210 Cks1b both reside on chromosome 1q21 and have been implicated in the pathogenesis of
211 gain1q MM (28). Simulating the upregulation of each gene individually and in combination
212 revealed a dose-dependent increase in cell numbers as the number of copies of genes on 1q
213 increased (Fig. 2E). Retrospective analysis of clinical data found a dose-dependent worsening
214 prognosis with increasing copies of chromosome 1q (29) (Fig. 2F). Strikingly, with a 5-fold
215 increase in expression of *MCL1* and *CKS1B* the cell population continues to increase beyond
216 the presence of any input signal (IKK activity returns to basal by 72 hours, Fig. S6). These
217 simulations indicated that amplification of the chromosomal region containing *MCL1* and *CKS1B*
218 may result in constitutive signalling and microenvironmental independence, which is
219 circumstantially supported by the observation that 1q21 is amplified in 91% of MM cell lines
220 (30).

221 We noted that within the models we were using, *cMYC* and *BCL2* in DH lymphoma impact the
222 cell cycle and apoptotic regulatory networks respectively, while *BCL6* is situated within
223 molecular networks controlling terminal B cell differentiation (Fig. S5). In the context of gain1q
224 MM, *CKS1B* and *MCL1* over expression results in the perturbation of the cell cycle and
225 apoptosis respectively. Consistent with DH lymphoma, we found that when mutations
226 simultaneously impact the cell cycle and apoptosis they combined deleteriously in simulations
227 and conferred poor prognosis in clinical data.

228 A computational pipeline enables simulations of heterogeneous patients from WES data
229 by mapping mutations to simulation parameters.

230 Having demonstrated the model can accurately predict known poor prognosis mutation
231 combinations, we next sought to establish whether other co-occurring deleterious mutations that

232 impact both the cell cycle and apoptosis could be identified through computational modelling.
233 Comparing simulations of activating mutations in *BCL2* and *cMYC* to simulations of an upstream
234 NEMO:IKK/NF-κB–activating mutation demonstrated that NF-κB-activating mutations could
235 cause an increase in both BCL2 and cMYC mRNA similar to that seen in double hit mutation
236 lymphomas (Fig. 3A). To employ the model to identify additional combinations of mutations that
237 could result in simultaneous anti-apoptotic and pro-proliferative molecular signalling, we sought
238 to leverage a WES dataset that captures the mutational heterogeneity in individual DLBCL
239 patients (7). We developed a pipeline to map both individual gene-level mutations and
240 chromosomal arm-level copy number alterations to computational parameters in the multi-scale
241 model (Fig. 3B, see Materials and Methods). This process created 113 patient-specific models,
242 with parameters modified based on gene alterations described from analysis of WES data (Fig.
243 3B), while all other parameters remained the same (31).

244 Personalised patient simulations identify an anti-apoptotic and pro-proliferative
245 subgroup of patients with poor prognosis.

246 Analysing pro-proliferative and anti-apoptotic factors over time for individual patient simulations
247 revealed highly heterogeneous expression between patients, with most dynamic changes
248 occurring in the first 6 hours (Fig. 4A). Concentrations at 6 hours were chosen to stratify patients
249 as this was the earliest time point at which most transient dynamics had passed but cells had
250 not been removed from the simulation due to cell division or cell death. Furthermore,
251 concentrations at later time points correlated strongly with those at 6 hours ($R^2 > 0.97$ between
252 6h, 9h and 12h time-points for both pro-proliferative and anti-apoptotic factors, Fig. S7). Patients
253 with above or below mean average expression of pro-proliferative or anti-apoptotic factors alone
254 at 6 hours did not differ in their progression-free survival (Fig. 4B, $p>0.05$). But stratification of
255 patients based on the combination of these factors identified a group of patients (19% of the

256 cohort) for which both anti-apoptotic and pro-proliferative (AAPP) factors were upregulated (Fig.
257 4C, example shown in orange in Fig. 4A). Patients classified as AAPP using computational
258 simulations had significantly worse prognosis than DLCBL patients not classified as AAPP (Fig.
259 3D, median PFS 26 months compared to not reached, p=0.001). The hazard ratio for AAPP
260 patients was 2.767 (95% confidence interval: 1.259-6.080) when compared to patients not
261 assigned AAPP (95% confidence interval: 0.664-1.506)

262 Comparing stratification through computational modelling with clustering of WES data revealed
263 that computational modelling enabled the identification of a substantially worse prognosis
264 patient cluster (AAPP) than could be identified from mutations alone. The AAPP cluster had a
265 median PFS of just 26 months compared to 77 months in the poor prognosis clusters (clusters
266 2/3/5 = C2/C3/C5) identified in reference 7 (Fig. 4D and E). Comparing cell of origin (COO) and
267 genetic clustering in the entire patient cohort with AAPP patients showed that AAPP patients
268 spanned all COO classifications and genetic clusters, but AAPP patients were enriched in poor
269 prognosis COO (ABC) and genetic cluster 5 (Fig. 4F and G).

270 Importantly, AAPP patients identified by modelling could not be identified from mutational
271 burden alone without simulation, and did not differ in their age at diagnosis, IPI, CNS
272 involvement, mutation count, number of driver mutations, number of copy number alterations, or
273 ploidy (q=0.1-1.0 for all clinical metrics) compared to the rest of the cohort (Fig. S8).

274 Only one mutation, Myd88:L265P, was significantly enriched in the AAPP patients (false
275 discovery correction using Benjamini-Hochberg, q=0.00027). The computational modelling had
276 not simply re-identified MYD88:L265P patients as the presence of Myd88:L265P was not
277 sufficient to confer poor prognosis in this cohort (Fig S9A). However, we did observe bi-modal
278 progression free survival among MYD88:L265P patients (Fig S9A and B). Strikingly, among
279 MYD88:L265P patients, those identified as AAPP had significantly worse PFS (Median PFS

280 MYD88:L265P+AAPP = 13.08 months, MYD88:L265P+not_AAPP=NR, p=0.0081), and overall
281 survival (median OS PFS MYD88:L265P+AAPP = 26.8 months, MYD88:L265P+not_AAPP=NR,
282 p=0.0081)(Fig S9C and D). AAPP patients are a larger subgroup (19%) than patients with either
283 DH or triple hit lymphoma (<5%)(32). Surprisingly, computational modelling did not assign
284 patients with concurrent structural variants affecting MYC and BCL2 to the AAPP category.
285 However, the cohort only contained 3 such patients was therefore not powered to assess their
286 outcome. Furthermore, it is worth noting that all three patients had low IPI (≤1) and none of
287 these patients had progressed at the end of the study. This suggests that computational
288 modelling of patients can identify poor prognosis patients independent of double hit structural
289 variant status.

290 Combining personalised computational simulations with consensus clustering enables
291 prognostic stratification.

292 When comparing mutational clustering with model-based stratification we observed that the
293 model had greater power to identify poor prognosis patients while mutational clustering excelled
294 in identifying good-prognosis patients (C1 and C4, Fig. 4E, had substantially better outcomes
295 than non-AAPP patients Fig. 4D, “Others”). Therefore, we devised a patient-stratification
296 approach using results from both methods, ultimately assigning patients to 3 patient categories
297 (AAPP=19%, C2/C3/C5 not AAPP = 57%, C1/C4 not AAPP = 24%, Fig. 5A and B). PFS in
298 these three clusters was significantly different (p=0.001), with AAPP patients having dismal
299 prognosis (PFS 26 months), C2/C3/C5 not AAPP patients having intermediate prognosis (PFS
300 93 months), and C1/C4 not AAPP patients having good prognosis (with 10-year PFS > 80%)
301 (Fig. 5C). This stratification approach was also robust to the inclusion or exclusion of patients
302 that did not contain a single mutation that mapped to a model parameter (Fig. S10).

303

304 Discussion

305 Predicting which blood cancer patients will respond well to treatment is key to empowering
306 patients and clinicians in clinical decision making and may be required to design trials that
307 advance the standard of care in heterogeneous blood cancers. We found that leveraging
308 genomic sequencing data from DLBCL patients to create personalised simulations and
309 combining this analysis with mutational clustering could outperform either approach in isolation.
310 Molecular clustering of DLBCL by multiple groups has identified clusters of patients that differ in
311 the signalling pathways most impacted by mutations (7-9). A single mutation may impact
312 multiple signalling networks, and multiple mutations may overcome or exacerbate each other.
313 For example, gain of 19q increases BAX (pro-apoptotic) and CCNE1 (pro-proliferative), NFKBIB
314 (pro-apoptotic and anti-proliferative). Computational modelling enabled identification of patients
315 with co-occuring mutations that are simultaneously pro-proliferative and anti-apoptotic, which
316 could not be determined from mutational clustering alone. Importantly, this modelling approach
317 identified patients with worse prognosis (PFS of 26 months) than genetic clustering and even
318 artificial intelligence-based approaches informed by more data (gene expression data,
319 mutational data, age and sex, poor prognosis cluster PFS 43.23 months)(32). We believe this
320 highlights the importance of encoding molecular network knowledge to contextualise mutational
321 information (12, 33).

322 There is compelling evidence supporting the adoption of genetic sequencing at diagnosis of
323 DLBCL to determine prognosis (34, 35). All mutations and CNAs modelled in this study affected
324 recurrently mutated genes, which are profiled by targeted sequencing panels. As such, the data
325 required for modelling could be generated at low cost and from fixed diagnostic samples, and

326 potentially even a plasma sample (36). Furthermore, the recent development of alternatives to
327 R-CHOP is likely to motivate utilisation of genomic sequencing data to rationally assign
328 therapies. There is emerging evidence supporting the addition of polatuzumab vedotin for high
329 risk DLCBL, the addition of Bortezomib in ABC-DLBCL, or the use of dose-adjusted EPOCH
330 plus rituximab for patients expressing high levels of MYC and BCL2 (37-39). Factoring in the
331 decreasing costs of sequencing, it seems likely that molecular profiling will become standard
332 diagnostic practice in DLBCL, with the aim of precisely identifying patients that would benefit
333 from an alternative to R-CHOP. Here we show that computational modelling can improve the
334 ability to identify such patients without requiring additional data.

335 More work is required if modelling is to become a widely used tool for personalised medicine
336 approaches. While computational simulations identified poor prognosis patients (AAPP),
337 mutational clustering was still needed to identify good-prognosis patients (C1/C4). Here we
338 chose not to create new models, but rather to test the utility of established computational
339 models of B cells that have not previously been applied to lymphoma (10, 11). We performed no
340 parameter fitting, as these parameters have been accumulated and validated across multiple
341 cellular contexts and cell types (10, 12, 13, 33). We assume that parameters, other than those
342 affected by mutations, remain consistent between healthy B cells and cancerous B cells. We
343 also assumed that all mutations within patients have the same effect size, equivalent to one
344 extra copy or loss of copy of each gene. The reality will be more complex, and we expect that
345 identifying and quantifying the impact of each mutation on each gene would improve the utility of
346 the model. Increasing the scope of the simulations, such that more recurrent mutations can be
347 directly assigned to model parameters is also likely to improve the utility of this approach.
348 However, the current model is clearly able to identify a population of very poor prognostic
349 patients that would not have been identified using mutational clustering and COO alone.

350 Of note, the B cell receptor (BCR) and toll-like receptor (TLR) signalling pathways, although not
351 explicitly modelled here, converge on NEMO-IKK and are the target of new therapeutic
352 advances and recurrent mutations. Similarly, epigenetic regulators such as EZH2 are
353 recurrently mutated but not explicitly modelled. Further work is also required to build on this
354 model and simulate how treatment may perturb biological networks harbouring mutations. Such
355 approaches may provide mechanistic insight into the development of relapsed DLBCL and
356 provide insight into second-line treatments. Simulating the effect of targeted inhibitors in the
357 patient-specific simulations presented here may provide insights into which patients will respond
358 to alternatives to R-CHOP, however, limited clinical data is available to validate such
359 predictions.

360 Performing large-scale computational simulations can be computationally challenging and
361 require substantial computational resources given the size and complexity of the molecular
362 networks simulated here (194 equations, and 563 parameters). Previous work has applied
363 logical modelling (molecular components are discretised into high/medium/low) to B cell
364 lymphoma to overcome this challenge (40). Here, we found a continuous modelling approach
365 was able to identify many gene dose-dependent effects that could not be identified with logical
366 modelling, including how chromosomal gain and amplification confer distinct prognoses. Despite
367 the requirement for substantial computational calculation, we found that simulating just 6 hours
368 of a single B cell per patient (simulating 100 cells required ~30 minutes of real-word time on a
369 2.1 GHz CPU) enabled us to stratify patients. The ability to perform short simulations is enabled
370 by experimental results that demonstrate that B cell molecular network states are rapidly
371 established and reliably inherited across multiple generations of cell division (10). It is likely that
372 clinical implementations of computational modelling could provide additional insight based on
373 simulations without specialised hardware and without delaying treatment.

374 Beyond DLBCL and MM, the modelling methodology described here may have utility in other
375 tumour types when mutational data and comprehensive, experimentally validated simulations
376 are available. Many of the recurrent mutational events in DLBCL and MM are common cancer-
377 associated mutations. cMyc deregulation is involved in the development of a wide variety of
378 cancers (41), Bcl2 is dysregulated in many malignancies including breast cancer and gastric
379 carcinoma (42, 43), and NF-κB is implicated in numerous cancers (44). A mathematical model,
380 simulating 17 cancer types, estimated that the number of carcinogenic mutations (hits) varies
381 from two to eight depending on the cancer type (45). With the increasing availability of cancer
382 genome data and the evolution of computational methods to identify the mutational burden of
383 cancer patients, the approach described here is rapidly becoming feasible in numerous cancers.
384 We expect these approaches to be most impactful in mutationally heterogeneous cancers, such
385 as breast cancer, where heterogeneity challenges early diagnosis, treatment selection and
386 prognosis prediction (46).

387 The ultimate goal of computational systems biology approaches such as the one presented here
388 is to enable personalised medicine by using models to get the right drugs to the right patients.
389 Modern sequencing techniques provide an abundance of data, but we have yet to fully utilise it.
390 Computational models with the power to translate patient-specific mutation data into
391 personalised prognostic and treatment predictions may be required to enable truly personalised
392 medicine approaches.

393 **Acknowledgements**

394 The authors would like to thank Ulf Klein and Jana Wolf for critical discussions and Eleanor
395 Jayawant for providing critical input into the manuscript. Funding for SM: Leukaemia UK John
396 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 conceptualised the study, and EM, JJ and TC conceptualised the Fig. 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).

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. Sehn LH, Salles G. Diffuse Large B-Cell Lymphoma. N Engl J Med. 2021;384(9):842-58.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. Mitchell S, Tsui R, Tan ZC, Pack A, Hoffmann A. The NF-κB multidimer system model: A knowledge base to explore diverse biological contexts. Science Signaling. 2023;16(776):eabo2838.
13. Shokhirev MN, Almaden J, Davis-Turak J, Birnbaum HA, Russell TM, Vargas JA, Hoffmann A. A multi-scale approach reveals that NF-κB cRel enforces a B-cell decision to divide. Mol Syst Biol. 2015;11(1):783.
14. Cloete I, Smith VM, Jackson RA, Pepper A, Pepper C, Vogler M, et al. Computational modeling of DLBCL predicts response to BH3-mimetics. NPJ Systems Biology and Applications. 2023;9(1):23.
15. Mitchell S, Mercado EL, Adelaja A, Ho JQ, Cheng QJ, Ghosh G, Hoffmann A. An NFκB activity calculator to delineate signaling crosstalk: type I and II interferons enhance NFκB via distinct mechanisms. Frontiers in Immunology. 2019;10:1425.
16. Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: A fresh approach to numerical computing. SIAM review. 2017;59(1):65-98.
17. Rackauckas C, Nie Q. Differentialequations. jl–a performant and feature-rich ecosystem for solving differential equations in julia. Journal of open research software. 2017;5(1).
18. Loriaux PM, Tesler G, Hoffmann A. Characterizing the relationship between steady state and response using analytical expressions for the steady states of mass action models. PLoS computational biology. 2013;9(2):e1002901.

457 19. Hawkins ED, Turner ML, Dowling MR, Van Gend C, Hodgkin PD. A model of immune regulation
458 as a consequence of randomized lymphocyte division and death times. Proceedings of the National
459 Academy of Sciences. 2007;104(12):5032-7.
460 20. Youle RJ, Strasser A. The BCL-2 protein family: opposing activities that mediate cell death. Nat
461 Rev Mol Cell Biol. 2008;9(1):47-59.
462 21. Heinzel S, Binh Giang T, Kan A, Marchingo JM, Lye BK, Corcoran LM, Hodgkin PD. A Myc-
463 dependent division timer complements a cell-death timer to regulate T cell and B cell responses. Nat
464 Immunol. 2017;18(1):96-103.
465 22. Dowling MR, Kan A, Heinzel S, Zhou JH, Marchingo JM, Wellard CJ, et al. Stretched cell cycle
466 model for proliferating lymphocytes. Proc Natl Acad Sci U S A. 2014;111(17):6377-82.
467 23. Cai C, Hu X, Dai P, Zhang T, Jiang M, Wang L, et al. c-Myc regulates neural stem cell quiescence
468 and activation by coordinating the cell cycle and mitochondrial remodeling. Signal Transduct Target
469 Ther. 2021;6(1):306.
470 24. Cheng Y, Luo H, Izzo F, Pickering BF, Nguyen D, Myers R, et al. m(6)A RNA Methylation Maintains
471 Hematopoietic Stem Cell Identity and Symmetric Commitment. Cell Rep. 2019;28(7):1703-16.e6.
472 25. Afifi MM, Crncec A, Cornwell JA, Cataisson C, Paul D, Ghorab LM, et al. Irreversible cell cycle exit
473 associated with senescence is mediated by constitutive MYC degradation. Cell Reports. 2023;42(9).
474 26. Meriranta L, Pasanen A, Alkodsi A, Haukka J, Karjalainen-Lindsberg M-L, Leppä S. Molecular
475 background delineates outcome of double protein expressor diffuse large B-cell lymphoma. Blood
476 advances. 2020;4(15):3742-53.
477 27. Schmidt TM, Fonseca R, Usmani SZ. Chromosome 1q21 abnormalities in multiple myeloma.
478 Blood Cancer J. 2021;11(4):83.
479 28. Hanamura I. Gain/Amplification of Chromosome Arm 1q21 in Multiple Myeloma. Cancers
480 (Basel). 2021;13(2).
481 29. Schmidt TM, Barwick BG, Joseph N, Heffner LT, Hofmeister CC, Bernal L, et al. Gain of
482 Chromosome 1q is associated with early progression in multiple myeloma patients treated with
483 lenalidomide, bortezomib, and dexamethasone. Blood Cancer Journal. 2019;9(12):94.
484 30. Hanamura I, Stewart JP, Huang Y, Zhan F, Santra M, Sawyer JR, et al. Frequent gain of
485 chromosome band 1q21 in plasma-cell dyscrasias detected by fluorescence in situ hybridization:
486 incidence increases from MGUS to relapsed myeloma and is related to prognosis and disease
487 progression following tandem stem-cell transplantation. Blood. 2006;108(5):1724-32.
488 31. Chakravarty D, Gao J, Phillips S, Kundra R, Zhang H, Wang J, et al. OncoKB: a precision oncology
489 knowledge base. JCO precision oncology. 2017;1:1-16.
490 32. Xu-Monette ZY, Zhang H, Zhu F, Tzankov A, Bhagat G, Visco C, et al. A refined cell-of-origin
491 classifier with targeted NGS and artificial intelligence shows robust predictive value in DLBCL. Blood
492 advances. 2020;4(14):3391-404.
493 33. Hoffmann A, Levchenko A, Scott ML, Baltimore D. The IkappaB-NF-kappaB signaling module:
494 temporal control and selective gene activation. Science. 2002;298(5596):1241-5.
495 34. Wright GW, Phelan JD, Coulibaly ZA, Roulland S, Young RM, Wang JQ, et al. A probabilistic
496 classification tool for genetic subtypes of diffuse large B cell lymphoma with therapeutic implications.
497 Cancer cell. 2020;37(4):551-68. e14.
498 35. Morin RD, Arthur SE, Hodson DJ. Molecular profiling in diffuse large B-cell lymphoma: why so
499 many types of subtypes? British Journal of Haematology. 2022;196(4):814-29.
500 36. Scherer F, Kurtz DM, Newman AM, Stehr H, Craig AF, Esfahani MS, et al. Distinct biological
501 subtypes and patterns of genome evolution in lymphoma revealed by circulating tumor DNA. Science
502 translational medicine. 2016;8(364):364ra155-364ra155.
503 37. Davies AJ, Barrans S, Stanton L, Caddy J, Wilding S, Saunders G, et al. Differential Efficacy From
504 the Addition of Bortezomib to R-CHOP in Diffuse Large B-Cell Lymphoma According to the Molecular

505 Subgroup in the REMoDL-B Study With a 5-Year Follow-Up. Journal of Clinical Oncology.
506 2023;41(15):2718-23.
507 38. Dodero A, Guidetti A, Tucci A, Barretta F, Novo M, Devizzi L, et al. Dose-adjusted EPOCH plus
508 rituximab improves the clinical outcome of young patients affected by double expressor diffuse large B-
509 cell lymphoma. Leukemia. 2019;33(4):1047-51.
510 39. Tilly H, Morschhauser F, Sehn LH, Friedberg JW, Trněný M, Sharman JP, et al. Polatuzumab
511 vedotin in previously untreated diffuse large B-cell lymphoma. New England Journal of Medicine.
512 2022;386(4):351-63.
513 40. Thobe K, Konrath F, Chapuy B, Wolf J. Patient-Specific Modeling of Diffuse Large B-Cell
514 Lymphoma. Biomedicines. 2021;9(11).
515 41. Kalkat M, De Melo J, Hickman KA, Lourenco C, Redel C, Resetca D, et al. MYC Deregulation in
516 Primary Human Cancers. Genes (Basel). 2017;8(6).
517 42. Joensuu H, Pylkkänen L, Toikkanen S. Bcl-2 protein expression and long-term survival in breast
518 cancer. Am J Pathol. 1994;145(5):1191-8.
519 43. Inada T, Kikuyama S, Ichikawa A, Igarashi S, Ogata Y. Bcl-2 expression as a prognostic factor of
520 survival of gastric carcinoma. Anticancer Res. 1998;18(3b):2003-10.
521 44. Luo JL, Kamata H, Karin M. IKK/NF-kappaB signaling: balancing life and death--a new approach
522 to cancer therapy. J Clin Invest. 2005;115(10):2625-32.
523 45. Anandakrishnan R, Varghese RT, Kinney NA, Garner HR. Estimating the number of genetic
524 mutations (hits) required for carcinogenesis based on the distribution of somatic mutations. PLoS
525 Comput Biol. 2019;15(3):e1006881.
526 46. Guo L, Kong D, Liu J, Zhan L, Luo L, Zheng W, et al. Breast cancer heterogeneity and its
527 implication in personalized precision therapy. Exp Hematol Oncol. 2023;12(1):3.

528

Figure Legends

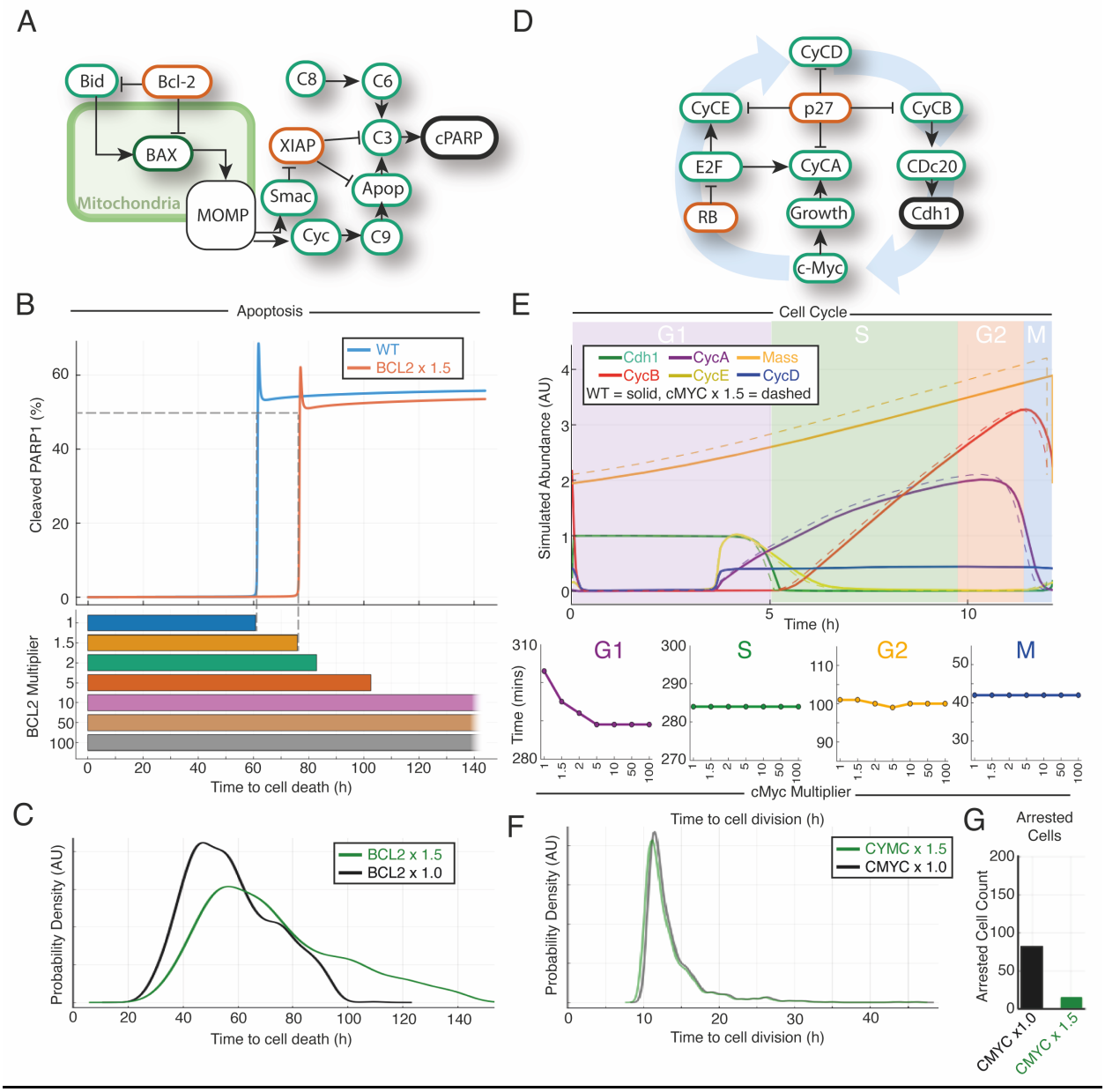

Fig. 1. Computational modelling of the cell cycle and apoptosis reveal limited impact of archetypal double hit mutations on their respective molecular networks. A-C Apoptosis model simulations**. D-F** Cell cycle model simulations. **A** Schematic of the apoptosis model leading to cleavage of PARP. **B** (top) The percentage of cleaved PARP over time for two

simulations (WT and BCL2 overexpressed) using the apoptosis model. There is a delay in PARP cleavage in the presence of overexpressed BCL2. **B** (bottom) Increasing time to death for cells with increasing BCL2 expression. **C** Graph showing distribution of time to death in simulation of 1000 cells, unmutated (black) compared to 1.5 fold BCL2 upregulation (green). **D** Schematic of the impact of cMyc on the cell cycle model. **E** (top) Output of the abundancies of different cell cycle proteins run to a limit cycle for a WT cell (solid line) and a cell in which cMYC is overexpressed (dotted line) showing a slight shortening of the cell cycle for some. CycA/D/E = Cyclin A/D/E, Cdh1= cdc20 homolog 1. **E** (bottom) Time for completion of cell cycle phases for cells with varying cMYC expression showing the main effect is in G1. **F** Distribution of time to cell division for simulations of heterogeneous populations of 1000 cells. Unmutated (black) compared to 1.5 fold cMYC upregulation (green) show very little difference. **G** Number of cells, from the simulations in panel F, for which the cell cycle has arrested in unmutated cells and cells in which cMYC is upregulated 1.5 fold. Upregulated cMyc substantially reduces the number of cells in cell cycle arrest.

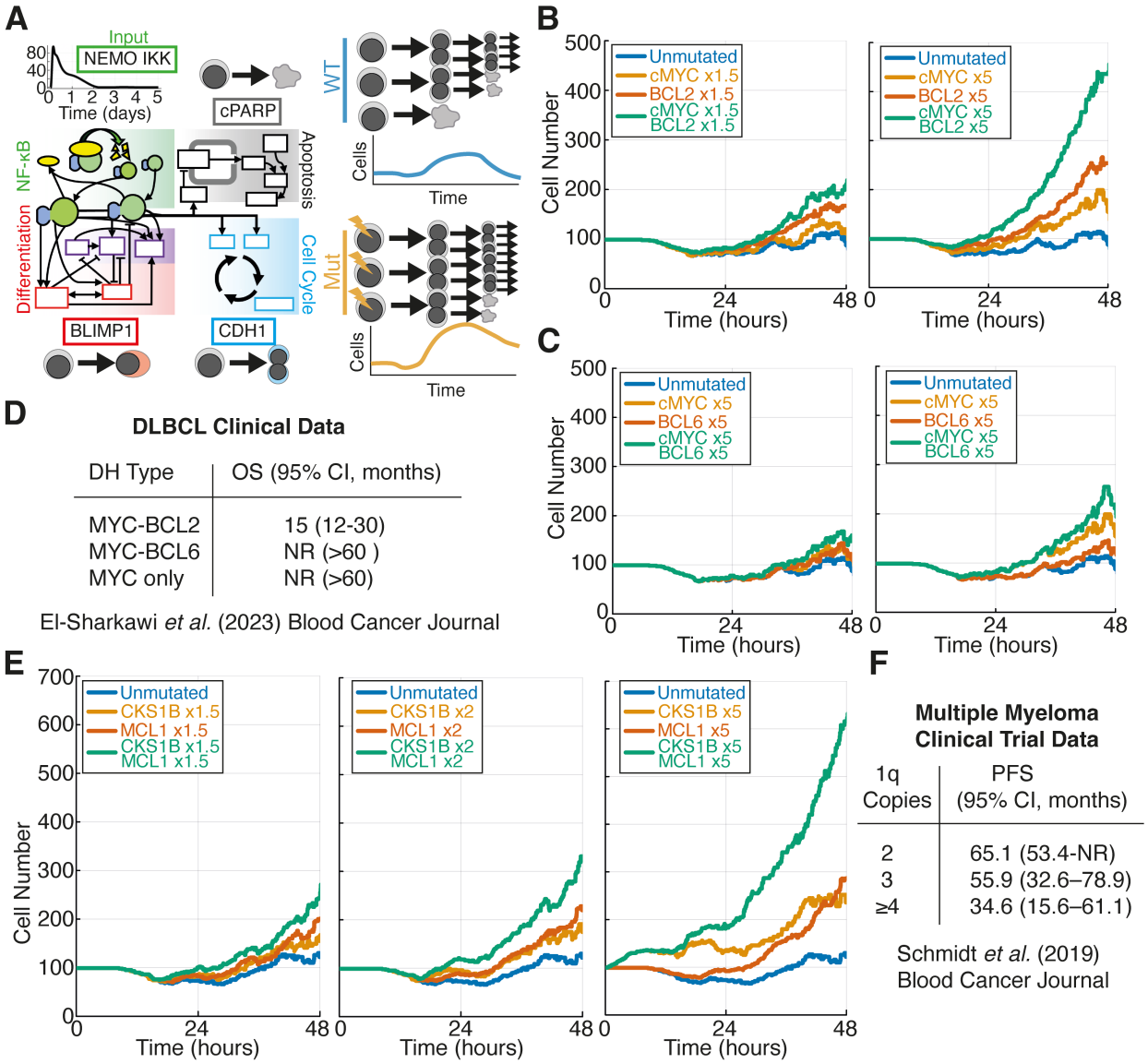

Fig. 2. Multi-scale modelling of mutations in subsets of DLBCL and Multiple Myeloma recapitulates clinical trial data. **A (left)** Simplified schematic model divided into component signalling pathways featuring NFκB signalling, apoptosis, the cell cycle and differentiation (adapted from ref 11), full detail provided in Supplementary Material. **A (right)** Schematic representation of the progression of cell lineages over time during multiscale simulations. The example depicts a mutation that allows cells to continue to proliferate when they would

otherwise die. **B&C** Simulated cell population size (cell count) over time for wild type (blue), individual mutations (orange and red) and double hits where both mutations are combined (green). **D** Overall survival (OS) data for groups of patients with MYC and BCL2/BCL6 mutations. **E** Cell population size (cell count) over time for simulations of gain1q multiple myeloma. CKS1B and MCL1 are located within chromosome 1q and therefore amplifications in this chromosome increase the abundance of both of these genes. **F** Progression-free survival (PFS) data for groups of patients with upregulation of CKS1B and MCL1 due to 2, 3 or 4 copies of chromosome 1q. NR = not reached.

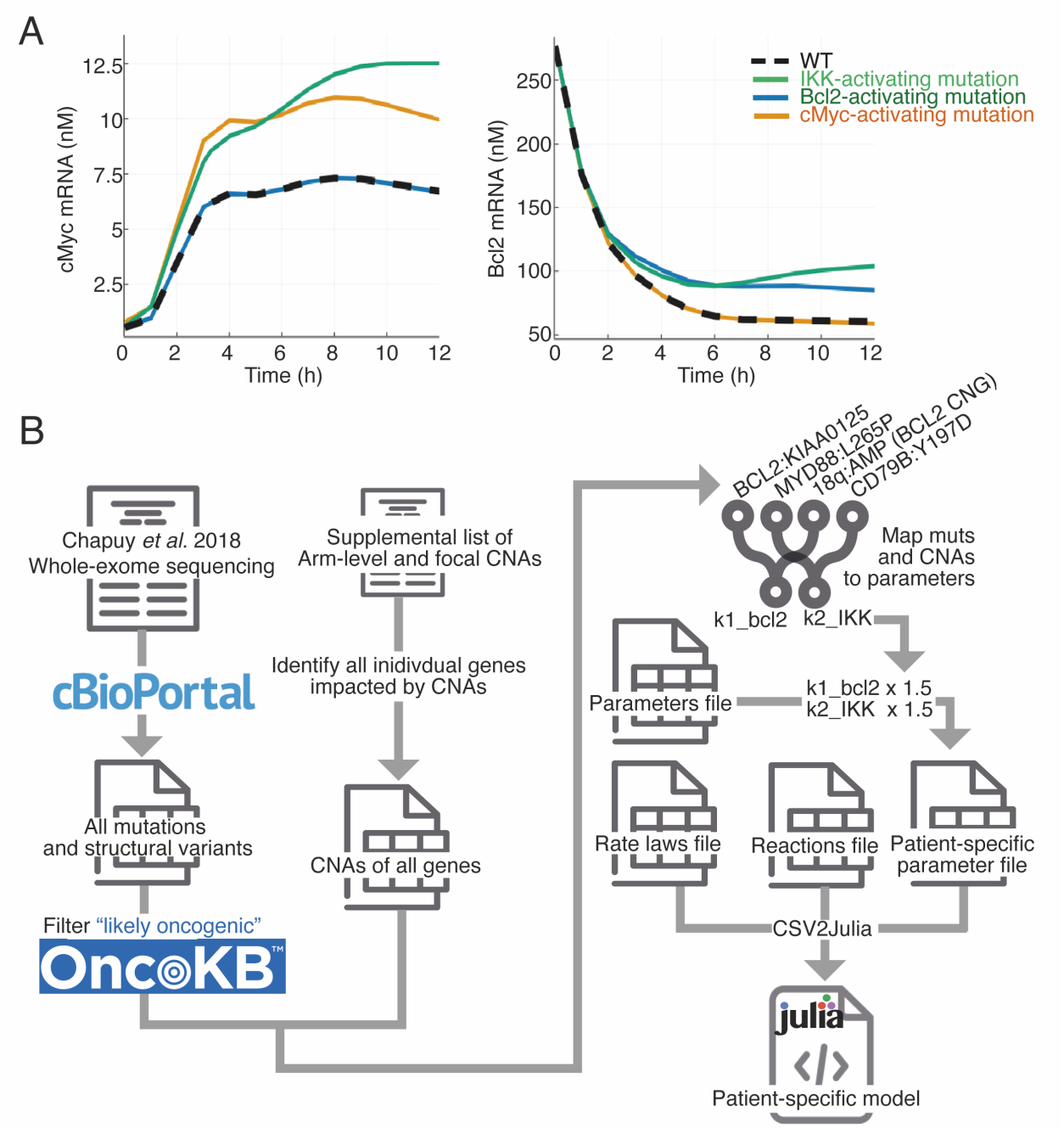

Fig. 3. Multiple mutations can create anti-apoptotic and pro-proliferative signalling. A Abundance of cMyc (A) and Bcl2 (B) mRNA in simulations of WT (dash), an IKK-activating mutation (green), a Bcl2-activating mutation (blue) and a cMyc-activating mutation (orange) over time. **B** Pipeline to incorporate mutational events from genetic sequencing to create patient

specific models. Example mutational mappings are shown including the model parameters they modify. The full mapping is provided on in the Github repository (https://github.com/SiFTW/norrisEtAl/blob/main/muts2Params.csv).

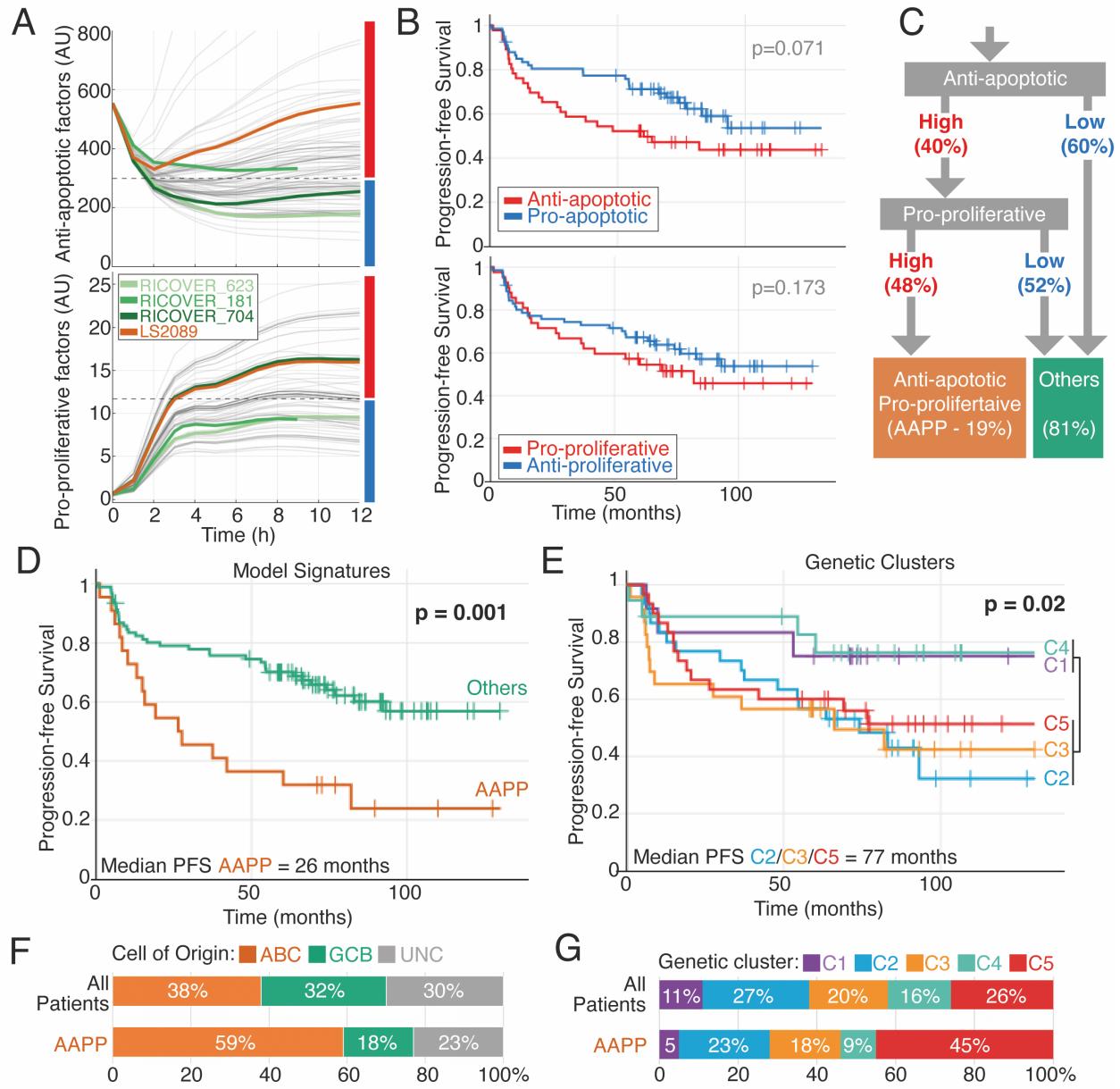

Fig. 4. Patient-specific modelling and stratification by pro-proliferation and anti-apoptotic species identifies poor prognosis patients within all cell-of-origin and genetic clusters. A Abundance of anti-apoptotic and pro-proliferative factors for individual patients (grey) simulated using the pipline in Fig 3. Example patients, labelled with their anonymised identifier, in each of 4 groups defined by the relative abundance of these factors are highlighted: low/low = light green; high anti-apoptotic = mid green; high pro-proliferative = dark green; anti-apoptotic and

pro-proliferative (AAPP) = orange. The dashed line indicates the mean value at 6h and is used for stratification into high and low in subsequent panels. **B** Kaplan-Meier (KM) plots comparing progression-free survival for DLBCL patients with model-identified high or low anti-apoptotic factors (top) and high or low proliferative factors (bottom). **C** Flow chart describing the identification of anti-apoptotic pro-proliferative (AAPP) patients. **D** KM plot for AAPP patients and all other patients. **E** KM plot for all modelled DLBCL patients grouped according to mutational clusters from ref. 7. **F&G** Grouped bar plots showing the percentage of patients in the AAPP group vs all patients, stratified by cell of origin **(F)** and genetic clusters **(G)**. ABC = Acitvated B Cell, GCB = Germinal Center B cell, UNC = unclassified.

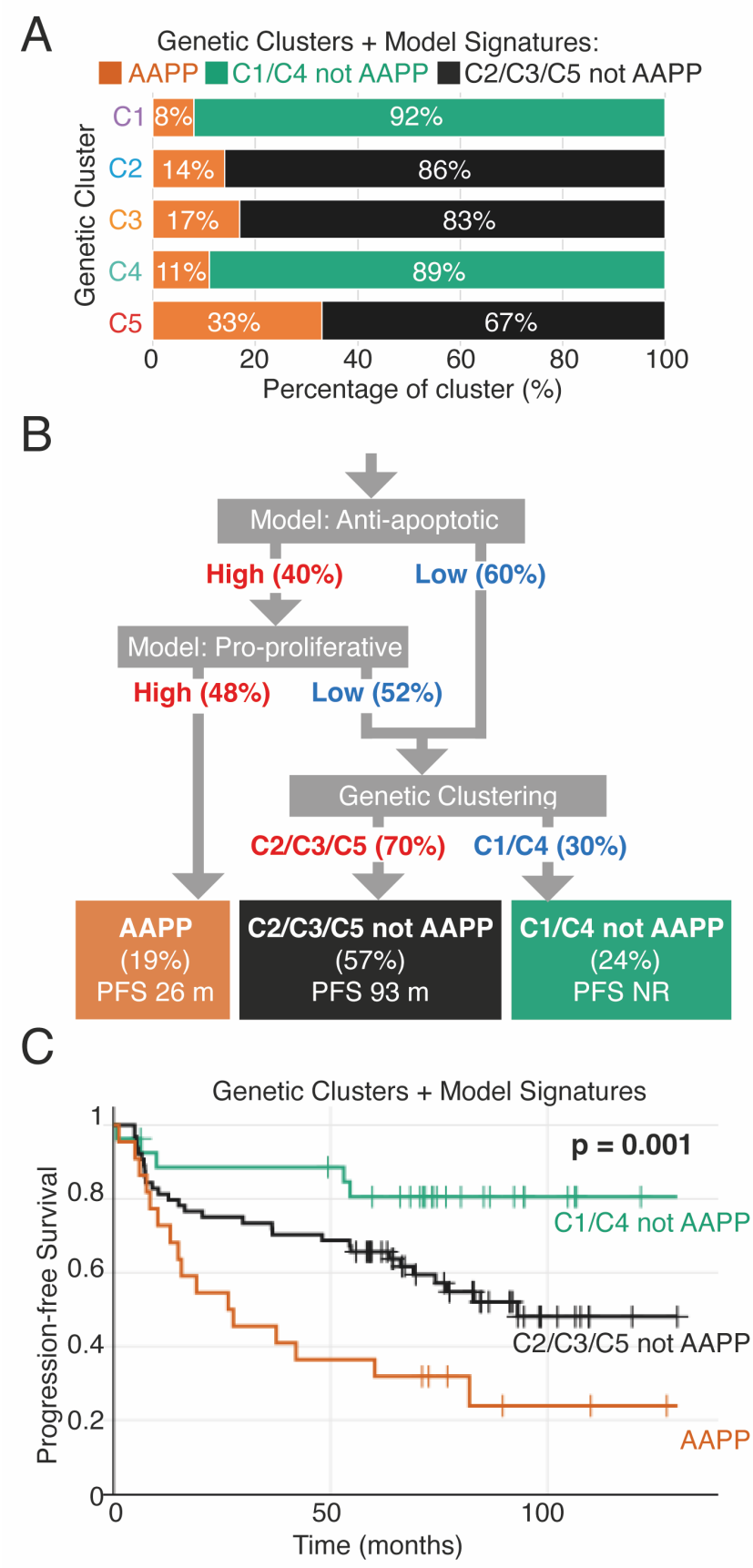
A
Genetic Clusters + Model Signatures:
AAPP
C1/C4 not AAPP
C2/C3/C5 not AAPP
Genetic Cluster
C1
8%
92%
C2
14%
86%
C3
17%
83%
C4
11%
89%
C5
33%
67%
0
20
40
60
80
100
Percentage of cluster (%)
B
Model: Anti-apoptotic
High (40%)
Low (60%)
Model: Pro-proliferative
High (48%)
Low (52%)
Genetic Clustering
C2/C3/C5 (70%)
C1/C4 (30%)
AAPP
(19%)
PFS 26 m
C2/C3/C5 not AAPP
(57%)
PFS 93 m
C1/C4 not AAPP
(24%)
PFS NR
C
Genetic Clusters + Model Signatures
p = 0.001
Progression-free Survival
1
0.8
0.6
0.4
0.2
0
C1/C4 not AAPP
C2/C3/C5 not AAPP
AAPP
0
50
100
Time (months)

Fig. 5. Combining patient-specific modelling with genetic clustering stratifies patients into distinct prognostic clusters.

A Grouped bar plot showing the percentage of AAPP and non-AAPP patients in each genetic cluster from ref. 7. **B** Flow chart describing stratification of patients first using patient-specific models to identify AAPP patients and then using clustering to stratify non-AAPP patients. **C** Kaplan-Meier plot of patients stratified using flow chart in panel B.