

Growth dynamics in naturally progressing chronic lymphocytic leukaemia

Michaela Gruber^{1,2,3,16,17}, Ivana Bozic^{4,17}, Ignaty Leshchiner^{2,17}, Dimitri Livitz^{2,17}, Kristen Stevenson⁵, Laura Rassenti⁶, Daniel Rosebrock², Amaro Taylor-Weiner², Oriol Olive¹, Reaha Goyetche¹, Stacey M. Fernandes¹, Jing Sun¹, Chip Stewart², Alicia Wong², Carrie Cibulskis², Wandi Zhang¹, Johannes G. Reiter⁷, Jeffrey M. Gerold⁷, John G. Gribben⁸, Kanti R. Rai⁹, Michael J. Keating¹⁰, Jennifer R. Brown^{1,11,12}, Donna Neuberg⁵, Thomas J. Kipps⁶, Martin A. Nowak^{7,13}, Gad Getz^{2,12,14,15,18*} & Catherine J. Wu^{1,2,11,12,18*}

How the genomic features of a patient's cancer relate to individual disease kinetics remains poorly understood. Here we used the indolent growth dynamics of chronic lymphocytic leukaemia (CLL) to analyse the growth rates and corresponding genomic patterns of leukaemia cells from 107 patients with CLL, spanning decades-long disease courses. We found that CLL commonly demonstrates not only exponential expansion but also logistic growth, which is sigmoidal and reaches a certain steady-state level. Each growth pattern was associated with marked differences in genetic composition, the pace of disease progression and the extent of clonal evolution. In a subset of patients, whose serial samples underwent next-generation sequencing, we found that dynamic changes in the disease course of CLL were shaped by the genetic events that were already present in the early slow-growing stages. Finally, by analysing the growth rates of subclones compared with their parental clones, we quantified the growth advantage conferred by putative CLL drivers *in vivo*.

Although recent genomic characterizations of cancer have provided a wealth of knowledge, the linking of genetic features of individual cancers with their phenotype has been largely lacking. The rate and pattern of growth can vary among patients with cancer, and various growth patterns have been used to describe cancer expansion^{1–10}. The two most commonly used models are exponential unbounded growth and sigmoidal growth, which stabilizes at a specific carrying capacity (for example, logistic growth). However, analyses of cancer growth patterns have so far either relied on samples with few data points per tumour or come from *in vitro* or mouse studies.

CLL is an informative model system for studying natural cancer growth, as it has a highly variable but often relatively indolent course, with treatment commonly withheld for months to decades until clinically necessary¹¹. Temporally dense tumour measurements are feasible by serial venipuncture, which enables the analysis of growth dynamics. The availability of highly pure tumour samples has already yielded rich, unbiased genetic characterization, leading to our current detailed understanding of the genomic attributes in CLL^{12–16}. Here, we undertake an integrative analysis of genetic data, clinical information and growth dynamics, including quantification of the effect of cancer mutations on growth rates measured in serial samples collected from 21 patients before therapy. We then corroborated the results in 86 independent patients with CLL, improving our understanding of the relationship between genotypes and phenotypes in cancer. Thus, we established a methodology and quantitative framework to model cancer growth observed directly from human patients, helping to explain the variation in the clinical behaviour of CLL.

Patterns of growth and genetic evolution in CLL

We identified 21 patients with CLL with serial samples collected between diagnosis and the start of treatment (Supplementary Table 1a). The time to the initiation of treatment ranged from 2.1 to 15.6 years (median 5.7). The cohort was balanced for the strong prognostic marker of the mutational status of the *IGH* variable region (*IGHV*)¹⁷ (Supplementary Methods). To assess the overall growth of the leukaemia in each patient, we evaluated 4 to 83 (median 22) pre-treatment measurements of peripheral white blood cell (WBC) counts per patient, the vast majority of which reflect CLL cells. A subset of patients exhibited bounded growth, with WBC counts plateauing over time, which suggests that the leukaemia can display logistic-like behaviour, whereas another distinct subset clearly demonstrated an exponential-like growth pattern (Fig. 1a, b, Extended Data Fig. 1b). We therefore devised a unified Bayesian model based on a logistic pattern, as it can model both predominant growth behaviours. In this context, the posterior probabilities of the growth rate (*r*) and carrying capacity (*K*) were estimated using a Markov chain Monte Carlo (MCMC) Gibbs sampler (Supplementary Methods). The dynamics of patient WBC counts that resulted in carrying capacities confidently below $1,000 \times 10^9$ cells per litre were classified as exhibiting logistic (LOG) growth, whereas those confidently above $1,000 \times 10^9$ cells per litre were categorized as displaying exponential-like (EXP) growth. Those cases that could not be confidently attributed to either of the categories were classified as having indeterminate (IND) growth patterns.

Using this classification scheme, five patients exhibited LOG growth behaviour, with stabilization at estimated carrying capacities of 71×10^9 – 264×10^9 cells per litre (Supplementary Methods). In ten

¹Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. ²Broad Institute of MIT and Harvard, Cambridge, MA, USA. ³Department of Internal Medicine I, Division of Haematology and Haemostaseology, Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria. ⁴Department of Applied Mathematics, University of Washington, Seattle, WA, USA. ⁵Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, Boston, MA, USA. ⁶Department of Medicine, University of California at San Diego Moores Cancer Center, La Jolla, CA, USA. ⁷Program for Evolutionary Dynamics, Harvard University, Cambridge, MA, USA. ⁸Barts Cancer Institute, Queen Mary, University of London, London, UK. ⁹Hofstra North Shore-LIJ School of Medicine, Lake Success, NY, USA. ¹⁰Department of Leukemia, MD Anderson Cancer Center, Houston, TX, USA. ¹¹Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. ¹²Harvard Medical School, Boston, MA, USA. ¹³Department of Mathematics and Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. ¹⁴Department of Pathology, Massachusetts General Hospital, Boston, MA, USA. ¹⁵Center for Cancer Research, Massachusetts General Hospital, Boston, MA, USA. ¹⁶Present address: CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria. ¹⁷These authors contributed equally: Michaela Gruber, Ivana Bozic, Ignaty Leshchiner, Dimitri Livitz. ¹⁸These authors jointly supervised this work: Gad Getz, Catherine J. Wu. *e-mail: gadgetz@broadinstitute.org; cwu@partners.org

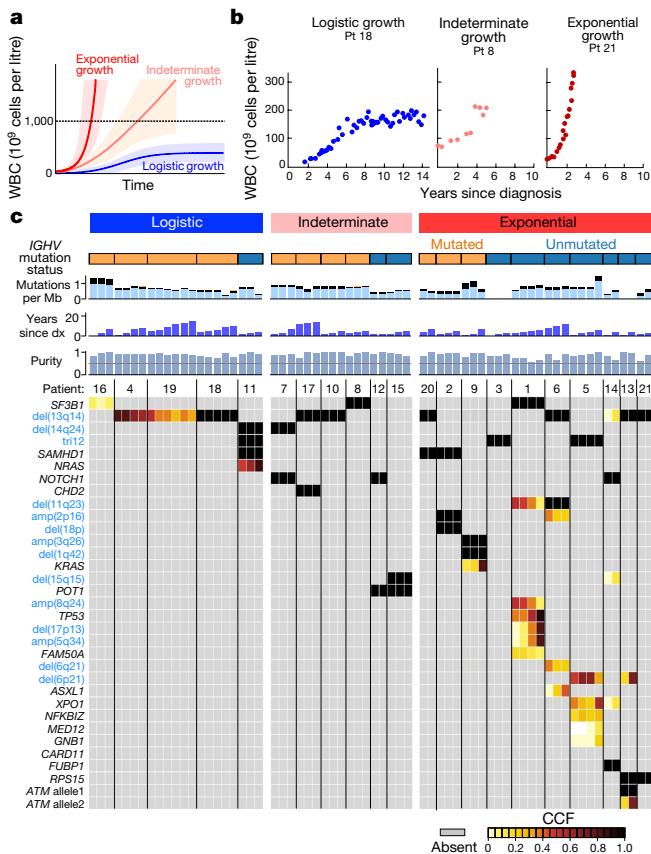


Fig. 1 | Growth dynamics and genetic changes in naturally progressing CLL. **a**, Schema of diverse growth patterns observable in CLL samples. **b**, Illustrative patient WBC dynamics over the period between diagnosis and the start of treatment. Pt, patient. **c**, Putative CLL drivers detected by WES analysis of pre-treatment serial samples from 21 patients, with assigned growth patterns, *IGHV* status. Cancer cell fractions (CCFs) of mutations are indicated. dx, diagnosis.

patients, the CLL cells exhibited EXP growth, with growth rates of 14–99% per year (median: 48% per year). Finally, in six patients, CLL growth could not be categorized with certainty as either LOG or EXP, either because of the short period of observation (patients 12 and 17) or owing to complex patterns of growth (patients 7, 8, 10 and 15) (Extended Data Fig. 1c, Supplementary Table 1). CLLs with LOG growth tended to have a longer time to treatment, with a median of 8.3 years (range 4.1–15.6) compared with 5.6 years (2.4–11.9) for EXP patients and 4.6 years (2.1–13.7) for the IND category; however, this difference was not significant across the three groups ($P = 0.19$, Kruskal–Wallis test; Supplementary Table 1a). Patients with LOG growth (4 out of 5) tended to have mutated *IGHV*—which is prognostically favourable—more frequently than patients with EXP growth (3 out of 10; $P = 0.12$; Fig. 1c).

To examine whether genetic differences underlie the distinct patterns of growth, we performed whole-exome sequencing (WES) analysis (median coverage of 107 \times) on a median of 3 (range 2–6) CLL samples per patient collected before receiving therapy, spanning 0.77–10.62 (median 3.28) years between the first and last available pre-treatment time points (Supplementary Tables 1b and 2). Consistent with previous studies^{12,13}, we detected a median overall somatic mutation frequency of 0.79 per megabase (Mb) across patients, with frequencies of 0.19 per Mb of silent events and 0.60 per Mb for non-silent somatic single nucleotide variants (sSNVs) and 0.031 per Mb for somatic insertions and deletions (sIndels) (Supplementary Table 3). In addition, we identified somatic copy number alterations directly from the WES data (Supplementary Table 4).

By comparing with previously identified putative CLL drivers^{12,13} (Supplementary Tables 5 and 6), we identified at least one driver in all

21 patients. We further examined whether the detected CLL drivers were clonal or subclonal by inferring their cancer cell fraction (CCF)¹⁵ (Supplementary Methods). The total number of driver mutations was lowest in patients whose CLLs demonstrated LOG growth (median 1, range 1–4) compared with those with IND growth (median 2, range 1–2), or EXP growth (median 4, range 2–7) ($P = 0.005$). We identified a trend towards a higher number of clonal drivers and an increased number of subclonal drivers in samples with EXP growth compared to those with LOG or IND growth ($P = 0.13$ and 0.019, respectively). Finally, we saw that, even without therapy, clones with CLL drivers, occurring in samples with EXP growth, exhibited larger shifts in their clonal fractions compared with clones in samples with IND and LOG growth ($P = 0.033$; Extended Data Fig. 2a, b). Most subclones with drivers present at the time of treatment were already detectable in the earliest samples.

Validation of global growth patterns

To confirm these findings, we studied an independent cohort of samples collected from 85 patients with previous characterization by WES, having at least four consecutive pre-treatment WBC measurements no more than two years apart^{14,15} (Supplementary Methods). Unlike the discovery cohort, patients in the extension cohort were not specifically selected to have undergone subsequent treatment. WES analysis was performed on samples collected at a median of 2.2 years (range 0, 18.9) after diagnosis. Of the 38 patients who subsequently underwent treatment, the median time from diagnosis to treatment was 2.6 years (range 0–18.9) (Supplementary Table 7). A median of 18 (range 4–73) pre-treatment measurements of circulating WBCs were available per patient.

As for the discovery cohort, we classified these 85 patients into LOG, EXP and IND growth categories (Supplementary Methods, Supplementary Table 8). Of 85 CLLs, 43 (51%) showed LOG growth (median estimated carrying capacity of 70×10^9 cells per litre, range 4–360), 12 showed EXP growth (14%; median growth rate 44, range 17–64%), and 30 had IND growth (35%; median growth rate 33%, range 4–197%) (Fig. 2a–c, Extended Data Figs. 2c, 3). Within CLLs categorized as having IND growth, 13 out of 30 (43%) had a short period of observation, 15 (50%) had a pattern of growth that appeared intermediate between LOG and EXP or was complex, and 2 (7%) had unevenly distributed data.

Consistent with the results in the discovery cohort, we again found that the mutated status of *IGHV* was most enriched in patients with LOG growth ($P = 0.002$). Similarly, we found a difference in the number of total, clonal and subclonal drivers among the groups, with larger values in patients with EXP growth ($P = 0.022$, $P = 0.011$ and $P = 0.017$, respectively; Fig. 2b, Supplementary Table 7a, b). Even among the 42 patients with known *IGHV* status in the LOG group, the 4 patients with unmutated *IGHV* had higher carrying capacities than the 38 patients with mutated *IGHV* (median of 189.2×10^9 versus 64.4×10^9 cells per litre; $P = 0.017$). We observed that samples with LOG growth had a higher proportion of patients that had either no known driver event or only del(13q) (51% LOG, 24% IND and 0% EXP; $P = 0.001$). In addition, the EXP group had a higher proportion of trisomy 12 than the IND and LOG cases (42% versus 10% and 9%, $P = 0.027$).

More patients with EXP and IND growth progressed to treatment than those with LOG growth (75%, 67% and 21%, $P < 0.001$) (Fig. 2d). We modelled the probability of treatment using univariable exact logistic regression and stepwise multivariable logistic regression. We considered age, sex, growth pattern (EXP versus LOG, IND versus LOG), *IGHV* status, fluorescence in situ hybridization cytogenetics, number of drivers, and the maximum and last available WBC count before treatment as potential covariates to assess their relative contribution to clinical outcome (Supplementary Methods). In univariable analysis, growth pattern, *IGHV* status, del(13q), last available WBC count and the number of total and clonal drivers were significant. The final multivariable model included growth rate pattern, del(13q) and last

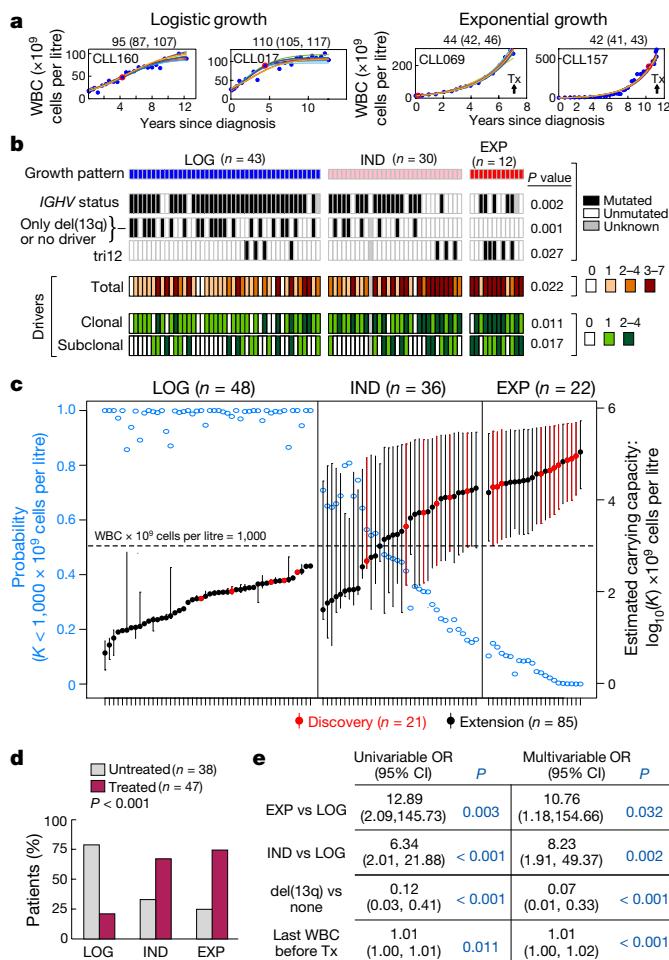


Fig. 2 | Integrative analysis of global CLL growth patterns and genetic attributes in an extension cohort. **a**, Examples of logistic (LOG; left) and exponential (EXP; right) growth patterns (70% credible intervals provided above each graph). Arrow indicates the start of treatment (Tx). Red dot denotes time of WES sampling. **b**, Associations between growth patterns and *IGHV* status, del(13q) or no driver, trisomy 12 and numbers of total, clonal and subclonal drivers. *P* values were determined by two-sided Fisher's exact test. **c**, Estimated carrying capacity (with 70% credible intervals, right axis) for patients with logistic, indeterminate (IND) and exponential growth, defined based on the posterior probability of carrying capacity (left axis), from discovery (red) and extension (black) cohorts. *n* indicates number of independent patients. **d**, Patients with exponential growth were more likely to require treatment. *P* values were determined by two-sided Fisher's exact test. **e**, Exact logistic regression modelling of the need for treatment relative to growth pattern for 79 patients having WES of a pre-treatment sample and fluorescence in situ hybridization cytogenetics data. CI, confidence intervals; OR, odds ratio. See Extended Data Table 1 for further details.

WBC count (Extended Data Table 1). Growth pattern contributed significantly (EXP versus LOG: *P* = 0.032; IND versus LOG: *P* = 0.002) to a model that already included del(13q) and the last available WBC count (Fig. 2e).

Growth patterns and evolution of relapse

Because clonal shifts were common in CLLs with more aggressive patterns of growth, we asked whether the growth pattern also affected clonal evolution after therapy. Of the 21 patients in the discovery cohort, 12 relapsed (median time from first until next treatment 1.88 years; range 0.52–5.56). Ten of these twelve patients had available post-treatment samples, which we also characterized by WES analysis.

Six of these ten patients had clear evidence of clonal evolution, characterized by a significant shift in the CCF (by 11–100%; Supplementary Methods) in at least one subclone (Fig. 3a, Extended

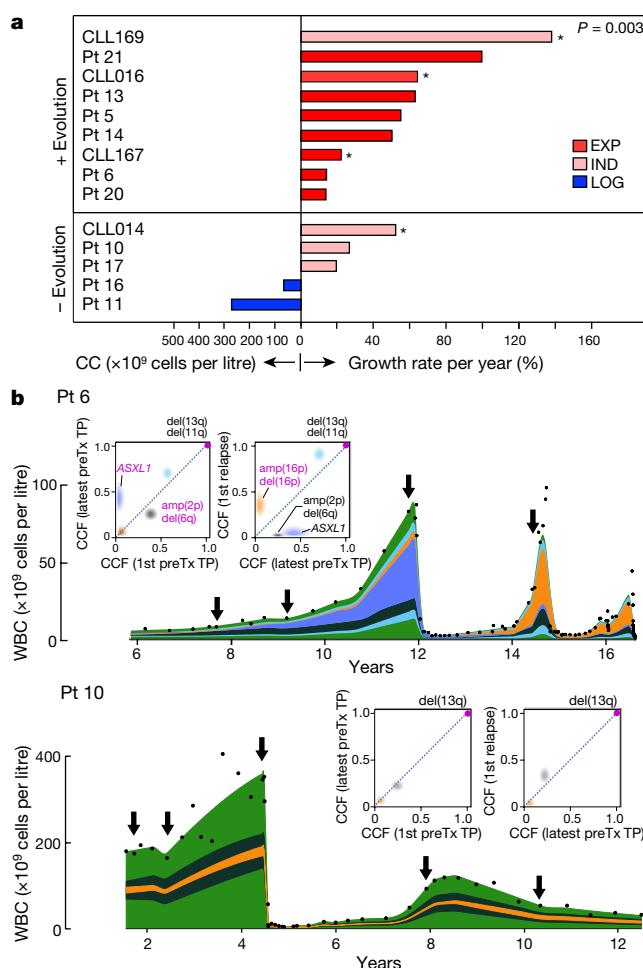


Fig. 3 | Treatment-associated evolution in relation to patterns of leukaemia growth. **a**, Right, pre-treatment growth pattern for 10 patients of the discovery cohort and 4 previously reported patients¹⁵ (asterisks) are related to the presence (top) or absence (bottom) of clonal evolution after relapse. Left, estimated carrying capacity (CC) for LOG patients. *P* values were determined by two-sided Fisher's exact test. **b**, Examples of integrated WBC dynamics (black dots indicate measurements) and genetic evolution between diagnosis and relapse for exponential growth and evolving subclones (patient 6), or indeterminate growth and inter-clonal balance (patient 10). Colours indicate subclones. Arrows indicate timing of WES samples. Insets show two-dimensional visualizations of changes in CCF across time points (TP).

Data Fig. 4a, Supplementary Table 9). In the other four patients, subclones retained interclonal balance even after cytoreductive treatment (Fig. 3a, Extended Data Fig. 4b). In two out of six evolving leukemias, we observed evidence of convergent evolution, with several branches affecting the same pathway. In patient 5, clone 5 with the mutation encoding GNB1(I80T) disappeared, but a distinct clone with GNB1(I80N) emerged together with a mutation in CDKN2A. In patient 13, one allele of ATM was clonally mutated and the second allele had a distinct subclonal mutation of ATM. After chemotherapy, the mutated ATM subclone was replaced by an expanding clone with del(11q) that encompassed ATM and BIRC3 on one allele, and two mutations in BIRC3 on the other. The four other patients with clonal evolution after relapse also revealed emergence of clones with driver events (Extended Data Fig. 4). All six patients with therapy-related clonal evolution of CLL exhibited EXP growth before treatment (growth rates of 14–99% per year). By contrast, patients with preserved clonal architecture despite therapy had CLL with either LOG or IND growth (with growth rates < 31% per year) (Fig. 3a, b, Extended Data Fig. 4). In patients from our previously reported study¹⁵, we found a similar association (Supplementary Methods, Supplementary Table 1c). Three of the

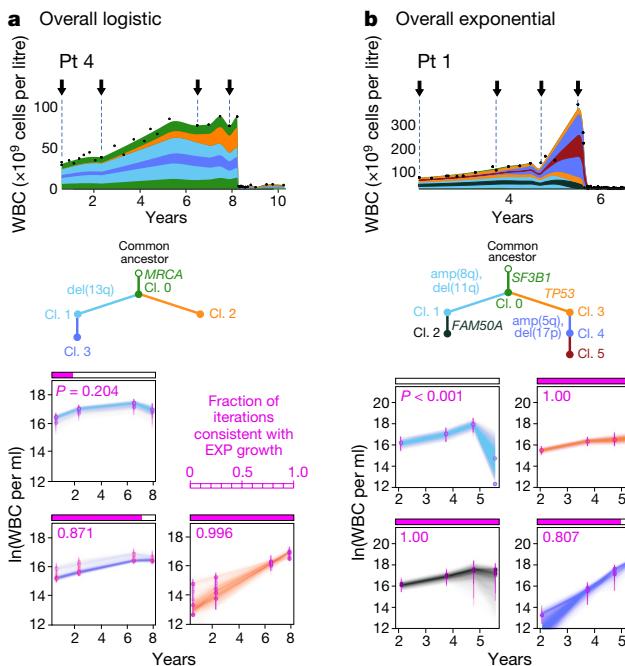


Fig. 4 | Subclonal growth patterns in untreated CLLs. **a, b,** Subclonal CLL growth patterns, estimated using PhylogeneticNDT, in overall logistic and exponential growth. P values represent significance of rejecting the null hypothesis of exponential growth, based on the proportion of Markov chain Monte Carlo (MCMC) iterations (magenta). The natural logarithms of numbers of cells in each clone (WBC per ml) are plotted against years from diagnosis. Error bars and trajectories represent uncertainties estimated by the MCMC. Further analysis of patients 5, 18 and 19 is shown in Extended Data Fig. 6b; and associated WGS analysis is shown in Extended Data Figs. 8, 9. Cl., clone.

four patients with data available to evaluate growth patterns (three patients included in the validation cohort of $n = 85$ plus one additional patient) had evolving clonal structure after relapse, and an IND or EXP growth pattern before treatment (growth rates of 32%, 64% and 138% per year), whereas the fourth patient with preserved clonal architecture despite therapy exhibited IND growth (Fig. 3, bars with asterisks). Overall, these results support the close association between the growth of leukaemia before treatment and clonal evolution after chemotherapy ($P = 0.003$).

Growth patterns of subclones

To identify whether distinct growth dynamics could be discerned among subclones underlying the overall growth patterns, we used information from macroscopic subclones (that is, having CCFs of $\geq 10\%$ at one or more time points) to derive probable phylogenies for each patient (Supplementary Methods, Supplementary Table 10, Extended Data Figs. 5–7). Using PhylogeneticNDT¹⁸, we identified subclones by n -dimensional clustering of CCFs of individual events across samples, estimated their phylogenetic relationships, and modelled the growth dynamics of each subclone (Supplementary Methods). We focused on the five patients with WES from four or more pre-treatment time points (patients 1, 4, 5, 18 and 19).

Of interest, even with overall LOG growth in the leukaemias of patients 4, 18 and 19, several subclones displayed behaviour consistent with EXP growth. Two out of four subclones (clones 1 and 3) of patient 4 had plateauing growth rates and even showed a decline in growth at the latest time point (Fig. 4a). By contrast, the growth of clone 2 was consistent with EXP growth, providing in vivo evidence of competition among subclones within an overall picture of LOG growth of the bulk tumour. It is possible that the growth pattern of this CLL would become EXP after takeover of the cancer cell population by clone 2. In the other two patients, most clones exhibited complex dynamics that did not

follow simple exponential or logistic patterns (Extended Data Fig. 6b). Clone 3 of patient 19 exhibited mild exponential growth over time. The other clones were mostly stable with fluctuations in their abundance. Thus, for cases with overall LOG growth, we observed that complex underlying intraclonal dynamics can result in net carrying capacity.

Patients 1 and 5, who had CLL with EXP growth, also revealed diverse interclonal dynamics (Fig. 4b, Extended Data Figs. 6, 7). Most subclones of patient 1 exhibited growth consistent with EXP patterns, with the later subclone (clone 4) probably showing faster exponential growth. In patient 5, the strongly growing subclones (clones 4 and 5) were only present in the last two time points, and thus we only tested clones 1–3 for behaviour consistent with EXP growth. Although the growth pattern of clone 1 did not fit an exponential pattern, both branches (clones 2 and 3) appeared to contribute to the overall EXP growth pattern of this case.

For added confidence and validation of the WES results, we also performed whole-genome sequencing (WGS) analysis of serial samples from a subset of cases with EXP (patients 1 and 6) and LOG (patient 4) growth patterns. The WGS results were concordant with those based on the WES data, while providing narrower credible intervals on the subclonal composition owing to an approximately 10-fold higher number of mutations per clone (Supplementary Methods, Extended Data Figs. 8, 9a–e).

Growth advantage of subclonal drivers

To quantify the growth advantage of individual subclones, we calculated the differences in growth rate (ΔGR) between child and parent clones (Supplementary Methods). We evaluated whether subclonal growth rates differed based on the presence or absence of known CLL drivers and if they had a significant selective advantage compared with the parent clone ($\text{Prob}[\Delta GR > 0] > 0.95$, which corresponds to $P < 0.05$) (Supplementary Table 11). We focused this analysis on the 35 subclones detected in 16 patients whose leukaemias displayed overall non-logistic WBC expansion (because those with overall LOG growth lacked abundant drivers). We inferred the distribution of growth rates of individual subclones and the differences to their parents using an MCMC-based method that samples an ensemble of likely phylogenetic trees for each patient (PhylogeneticNDT GrowthKinetics¹⁸; Supplementary Methods). The model also takes into account the number of reads supporting each somatic mutation, tumour purity, absolute copy number, and the WBC measurements (Extended Data Fig. 5d).

Indeed, 7 subclones from 6 patients (patients 1, 5, 6, 9, 13 and 14) contained a known CLL driver and their growth was significantly higher than their parent subclone (Fig. 5a, Extended Data Fig. 9f). The strongest accelerations were associated with second hits in the tumour suppressor genes *TP53* or *ATM* (ΔGR of 127% and 90% per year for clone 4 of patient 1 and clone 2 of patient 13, respectively). For patient 14, one subclone with several driver mutations (*XPO1*, *del(13q)*, *del(15q)*) also expanded more rapidly than its parent (ΔGR of 57% per year). The CLL of patient 5 showed high overall growth driven by the two independent subclones containing mutations in *MED12* and *GNB1* (ΔGR of 41% and 78% per year, respectively). We further saw strong growth acceleration in patient 6 (*ASXL1*, ΔGR of 82% per year) and in patient 9 (*KRAS*, ΔGR of 34% per year). These analyses provide an *in vivo* measurement and functional evidence of the cancer-driving capacity of these mutations. One subclone (from patient 7) had a growth rate indicative of a fitness advantage compared with its parent but without a known CLL driver (Extended Data Fig. 9g). Its estimated ΔGR was lower than that of subclones with known drivers (ΔGR of 16% ($\pm 10\%$ s.d.) per year). Another set of 8 subclones from 4 patients were identified to have mutations in known CLL drivers, but lacked a significant growth advantage compared with their parent subclone (for example, clone 2 in patient 3 with mutated *CARD11*; ΔGR 10%, $P = 0.20$; Extended Data Fig. 9h). Finally, more than half of the subclones (19 out of 35) from 12 patients appeared as ‘passenger’ subclones with no relative advantage compared with the parental clones

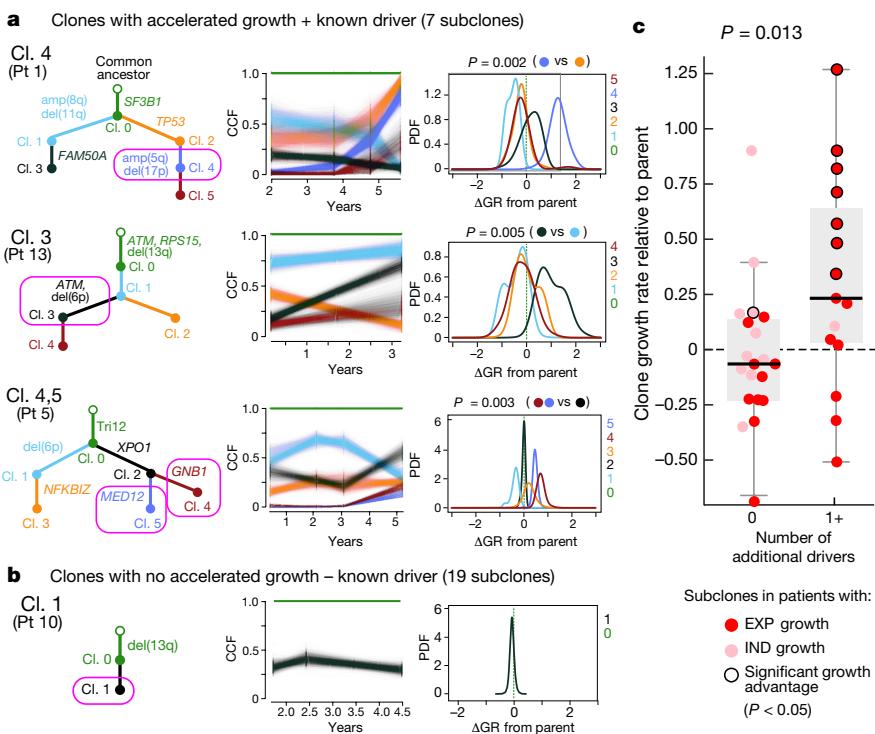


Fig. 5 | Selective growth advantage of subclones in CLL. a, b, Examples of subclones (magenta boxes) with significant growth advantage relative to their parent that contain known drivers (a) and a subclone without drivers and no growth acceleration (b). Results from PhylogeNNT analysis include: most likely phylogenetic tree (left); permutations of sSNVs during tree construction yielding posterior CCFs of the clusters (with 95%

and without any known CLL driver. These subclones possibly represent genetic drift^{19,20} or may have had an advantage for a short period of time (for example, clone 1 in patient 10, Fig. 5b).

In 7 out of 16 patients, we observed more than one class of subclonal growth simultaneously (Extended Data Fig. 7). Subclones with known CLL drivers were more likely to have accelerated growth relative to their parent (15 clones, median ΔGR of 0.23 (interquartile range (IQR) 0.02–0.71)) compared with subclones without additional known drivers (20 clones, with median ΔGR of −0.07 (IQR −0.23–0.13); $P = 0.013$, in a linear mixed model) (Fig. 5c).

Finally, patients 1, 5, 6, 9 and 13 CLL had subclones with accelerated growth and branched phylogenies, providing opportunities to examine the relative fitness advantage that is conferred by drivers to the subclone within specific genetic contexts. For example, the acquisition of amp(5q34) and del(17p) in clone 4 (patient 1) seems to confer faster growth than clones with the TP53 mutation alone (clone 2), del(11q) and amp(8q) (clone 1) or with the FAM50A mutation (clone 3) in the context of clonal SF3B1 mutation. Similarly, in patient 6, clone 3 had a mutation in ASXL1 in the context of clonal del(13q) and del(11p) and expanded faster than a sibling clone having amp(2p) and del(6q) mutations. With sufficiently large multi-time point sequencing datasets and overall tumour burden information (that is, WBC counts), comparing the growth of sibling clones will enable the construction of a map of the fitness of driver events in specific genetic contexts.

Discussion

Many mathematical models have been proposed to describe cancer growth in patients, but with no consensus^{7–9}. Previous works reported exponential growth in leukaemia, lung, liver and breast cancer^{7,8}, and sigmoidal growth (such as logistic or Gompertz) in earlier studies of breast cancer^{4,5} and in xenografts derived from human colorectal cancer⁶. Our study shows that individual cancers of the same type can exhibit diverse growth patterns: whereas many patients with CLL exhibited exponential-like growth with high or unbounded carrying capacity,

credible intervals) (middle); and growth rates relative to parental clones (right). Significance of differential growth rate ($\Delta GR > 0$) was estimated based on the MCMC. c, Linear mixed model for difference between clones without ($n = 20$) and with ($n = 15$) putative CLL drivers. Black circles denote significant growth advantage. Horizontal black line indicates median, and whiskers extend to the last datapoint within $1.5 \times$ the IQR.

in others, the disease achieved a—potentially temporary—stable state, and exhibited a logistic-like pattern with finite carrying capacity. For these distinct growth classes, we further identified clear differences in genetic attributes and clinical outcomes: exponential growth was associated with a larger number of CLL drivers, greater subclonal dynamics and an earlier need for treatment. Conversely, logistic growth was associated with a narrower spectrum of genetic alterations, fewer subclonal drivers, and interclonal stability even in relapse after treatment. We and others previously noted that del(13q) and trisomy 12 are probable early events in CLL phylogeny^{13,21}, and we found that the presence of del(13q) is associated with logistic growth, whereas trisomy 12 was associated with exponential growth. Our multivariable analysis demonstrated that the incorporation of growth pattern behaviour provides additional explanatory power above that provided by the presence of the strong prognostic cytogenetic marker of del(13q)²², and was stronger than the mutational status of *IGHV*¹⁷. Furthermore, the genetic diversity of CLL seems to be established early in the disease, consistent with previously reported smaller series^{23–25}. In the absence of therapeutic perturbations, the observed rate and pattern of outgrowth of CLL over time thus seems to ‘play out’ the pre-determined genetic and clonal composition. Of note, the growth rates of these clones from untreated patients are several-fold lower than that of relapse clones we previously reported in patients receiving ibrutinib². Recent studies to characterize genetically the premalignant states of solid tumours²⁶, and of monoclonal B cell lymphocytosis^{27,28} (considered a precursor condition of CLL), have further supported the idea that the extent of genetic diversity of malignancies such as CLL long predates its full-blown clinical picture.

Our approach of integrating patient tumour burden measurements and sequencing data from several disease time points provides a conceptual framework for understanding the growth trajectories of individual populations of CLL cells. We detected evidence of in vivo competition among subclones in early CLL and more atypical behaviours that could not be categorized using simple models, extending

previously reported insights on CLL evolution during later periods of the disease course^{29–31}. For example, in a subset of patients, we observed a shift from overall logistic growth, with many years spent at carrying capacity, to exponential growth.

Finally, our computational framework enabled us to quantify the degree of growth acceleration and fitness of genetically defined subclones over their parental clones. Accelerated growth was strongly enriched in subclones with well-established CLL drivers. Thus, we provide evidence of ongoing evolution *in vivo*, with clear growth-accelerating roles and potential synergies of driver mutations. We also find frequent existence of growth-neutral subclones without drivers, probably representing genetic drift^{19,20}. These fundamental findings are especially relevant to the ongoing efforts of precision oncology, in which the estimation of clonal growth dynamics in individual patients may inform therapy and predict the course of their disease.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All relevant data are available from the authors and/or are included with the manuscript. Clinical data about patients and samples analysed in the discovery cohort are listed in Supplementary Table 1a, b; sequencing metrics and somatic mutations are provided in Supplementary Tables 2–4. WES data are in dbGaP under accession code phs001431.v1.p. For the extension cohort, patient and sample characteristics as well as sequencing data are available from a previous publication¹⁵, and clinical data are summarized in Extended Data Table and Supplementary Tables 7 and 8. Further data to assess WBC dynamics were collected from these patients for this study and are illustrated in Extended Data Fig. 2, with clinical characteristics of patients with additional relapse samples provided separately in Supplementary Table 1c.

Code availability

Phylogenetic NDT package¹⁸ is available at <https://github.com/broadinstitute/PhylogeneticNDT>. Phylogenetic NDT uses Python 2.7.13 and the following Python modules available from pypi.org: bottle 0.12.13, dill 0.2.7.1, et-xmlfile 1.0.1, intervaltree 2.1.0, jsonschema 2.6.0, lxml 3.7.3, more-itertools 2.5.0, mpmath 0.19, networkx 1.11, openpyxl 2.4.1, pdfkit 0.6.1, pydotplus 2.0.2, pync 2.3.6, pycm3 3.0, python-dateutil 2.6.1, rpy2 2.8.5, seaborn 0.7.1, simplejson 3.10.0, svgwrite 1.1.9, scikit-learn 0.18.1, biopython 1.68. In addition, pyemd (<https://github.com/garyordanji/pyemd>) and sselogsumexp (<https://github.com/rmcgibbo/sselogsumexp>) modules were used. Code for the Bayesian modelling of growth patterns is available at: <https://github.com/ivbozic/Bayesian-Growth-Pattern-Modeling>.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests are available at <https://doi.org/10.1038/s41586-019-1252-x>.

Received: 26 August 2017; Accepted: 1 May 2019;

Published online 29 May 2019.

- Wodarz, D. & Komarova, N. L. *Dynamics of Cancer: Mathematical Foundations Of Oncology* (World Scientific, 2014).
- Burger, J. A. et al. Clonal evolution in patients with chronic lymphocytic leukaemia developing resistance to BTK inhibition. *Nat. Commun.* **7**, 11589 (2016).
- Diaz, L. A. Jr. et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537–540 (2012).
- Norton, L. A Gompertzian model of human breast cancer growth. *Cancer Res.* **48**, 7067–7071 (1988).
- Spratt, J. A., von Fournier, D., Spratt, J. S. & Weber, E. E. Decelerating growth and human breast cancer. *Cancer* **71**, 2013–2019 (1993).
- Misale, S. et al. Vertical suppression of the EGFR pathway prevents onset of resistance in colorectal cancers. *Nat. Commun.* **6**, 8305 (2015).
- Talkington, A. & Durrett, R. Estimating tumor growth rates *in vivo*. *Bull. Math. Biol.* **77**, 1934–1954 (2015).
- Rodriguez-Brenes, I. A., Komarova, N. L. & Wodarz, D. Tumor growth dynamics: insights into evolutionary processes. *Trends Ecol. Evol.* **28**, 597–604 (2013).
- Gerlee, P. The model muddle: in search of tumor growth laws. *Cancer Res.* **73**, 2407–2411 (2013).
- Hart, D., Shochat, E. & Agur, Z. The growth law of primary breast cancer as inferred from mammography screening trials data. *Br. J. Cancer* **78**, 382–387 (1998).
- Hallek, M. et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood* **131**, 2745–2760 (2018).
- Puente, X. S. et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* **526**, 519–524 (2015).
- Landau, D. A. et al. Mutations driving CLL and their evolution in progression and relapse. *Nature* **526**, 525–530 (2015).
- Wang, L. et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365**, 2497–2506 (2011).
- Landau, D. A. et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714–726 (2013).
- Döhner, H. et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N. Engl. J. Med.* **343**, 1910–1916 (2000).
- Hamblin, T. J., Davis, Z., Gardiner, A., Oscier, D. G. & Stevenson, F. K. Unmutated Ig V_H genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* **94**, 1848–1854 (1999).
- Leshchiner, I. et al. Comprehensive analysis of tumour initiation, spatial and temporal progression under multiple lines of treatment. Preprint at <https://www.biorxiv.org/content/10.1101/508127v2> (2018).
- Bozic, I., Gerold, J. M. & Nowak, M. A. Quantifying clonal and subclonal passenger mutations in cancer evolution. *PLOS Comput. Biol.* **12**, e1004731 (2016).
- Williams, M. J., Werner, B., Barnes, C. P., Graham, T. A. & Sottoriva, A. Identification of neutral tumor evolution across cancer types. *Nat. Genet.* **48**, 238–244 (2016).
- Wang, J. et al. Tumor evolutionary directed graphs and the history of chronic lymphocytic leukemia. *eLife* **3**, e02869 (2014).
- The International CLL-IPI working group. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncol.* **17**, 779–790 (2016).
- Ojha, J. et al. Deep sequencing identifies genetic heterogeneity and recurrent convergent evolution in chronic lymphocytic leukemia. *Blood* **125**, 492–498 (2015).
- Rose-Zerilli, M. J. J. et al. Longitudinal copy number, whole exome and targeted deep sequencing of ‘good risk’IGHV-mutated CLL patients with progressive disease. *Leukemia* **30**, 1301–1310 (2016).
- Smith, E. N. et al. Genetic and epigenetic profiling of CLL disease progression reveals limited somatic evolution and suggests a relationship to memory-cell development. *Blood Cancer J.* **5**, e303 (2015).
- Martinez, P. et al. Dynamic clonal equilibrium and predetermined cancer risk in Barrett’s oesophagus. *Nat. Commun.* **7**, 12158 (2016).
- Ojha, J. et al. Monoclonal B-cell lymphocytosis is characterized by mutations in CLL putative driver genes and clonal heterogeneity many years before disease progression. *Leukemia* **28**, 2395–2398 (2014).
- Barrio, S. et al. Genomic characterization of high-count MBL cases indicates that early detection of driver mutations and subclonal expansion are predictors of adverse clinical outcome. *Leukemia* **31**, 170–176 (2017).
- Schuh, A. et al. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood* **120**, 4191–4196 (2012).
- Braggio, E. et al. Longitudinal genome-wide analysis of patients with chronic lymphocytic leukemia reveals complex evolution of clonal architecture at disease progression and at the time of relapse. *Leukemia* **26**, 1698–1701 (2012).
- Amin, N. A. et al. A Quantitative analysis of subclonal and clonal gene mutations before and after therapy in chronic lymphocytic leukemia. *Clin. Cancer Res.* **22**, 4525–4535 (2016).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Acknowledgements We are grateful to P. dal Cin, D.-A. Landau, S. Shukla and U. Jäger for discussions. We also appreciate the efforts of all study nurses and clinical staff that made this study feasible, and the patients who generously provided their samples for this research. This work was supported in part by the NCI (5PO1CA081534-14, 1R01CA155010-01A1, P01CA206978, U10CA180861), the CLL Global Research Foundation, and by NHLBI (1R01HL103532-01). M.G. was supported by a Marie-Curie International Outgoing Fellowship from the European Union (PIOF-2013-624924). G.G. is partially supported by the Paul C. Zemecnik Chair in Oncology at the Massachusetts General Hospital Cancer Center. C.J.W. is a Scholar of the Leukemia and Lymphoma Society.

Author contributions M.G., I.B., I.L., D.L., D.N., M.A.N., G.G. and C.J.W. designed the study, analysed and interpreted data. I.B. modelled CLL growth patterns across patients. I.L., D.L., I.B. and G.G. developed methods for the analysis of genomic data and modelled the clonal structure, trees and growth dynamics of individual clones. I.L., D.L., M.G. and G.G. performed analysis of genomic data. M.G., L.R., S.M.F., O.O. and R.G. collected samples and annotations. J.G.G., K.R.R., M.J.K., J.R.B. and T.J.K. oversaw patient care. M.G., L.R., W.Z., A.W. and C.C. performed sample isolation and analysis. K.S. and D.N. performed statistical analysis. D.R., C.S., J.S., J.G.R., J.M.G. and A.T.-W. contributed to the analysis of genomic data. M.G., I.B., I.L., D.L., K.S., D.N., G.G. and C.J.W. wrote the manuscript. All authors read and approved the final manuscript.

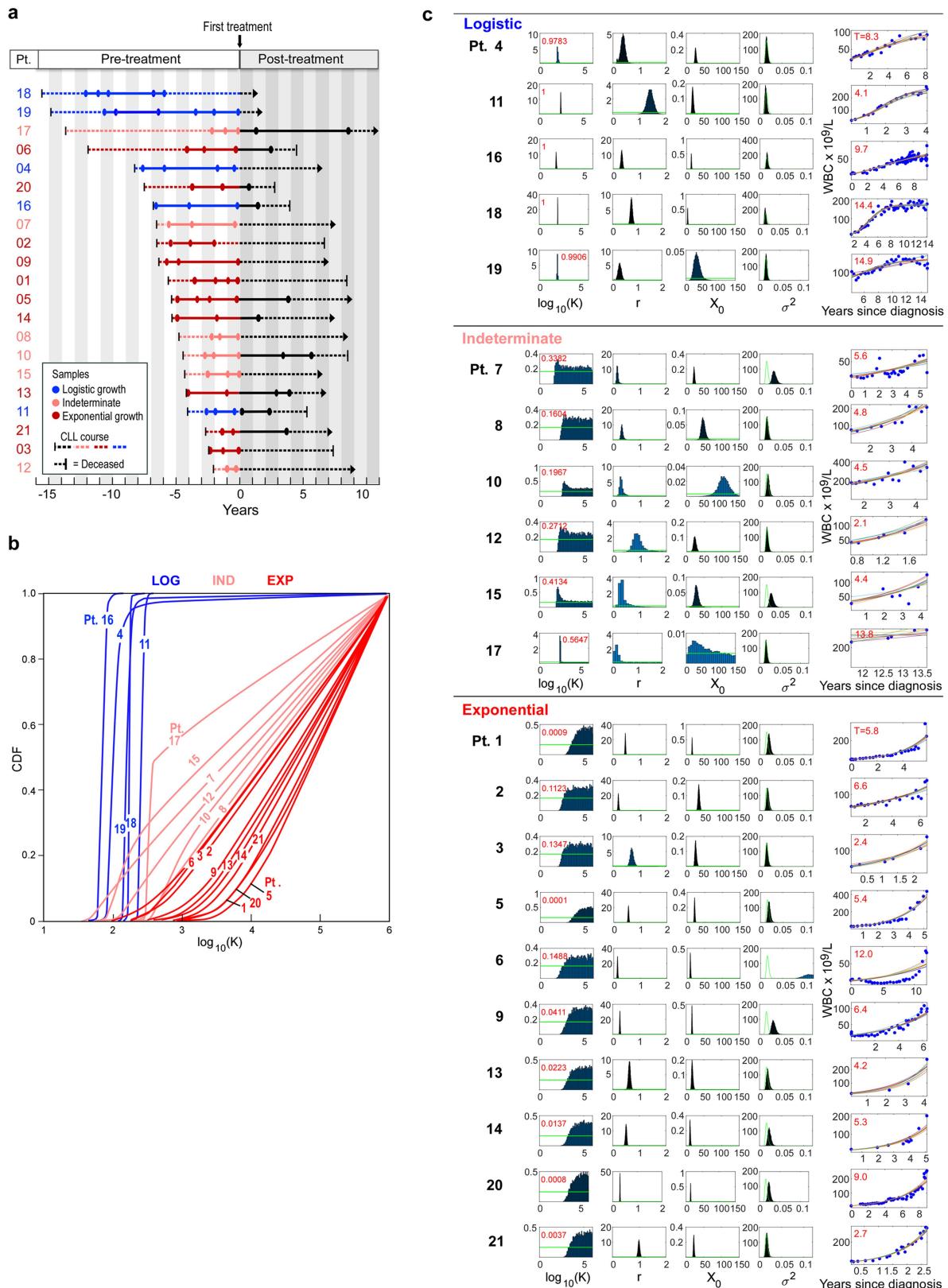
Competing interests C.J.W. is founder of Neon Therapeutics and a member of its scientific advisory board. G.G. receives research funds from IBM and Pharmacyclics. G.G. is an inventor of several bioinformatics-related patents, including patents related to MuTect and ABSOLUTE. C.J.W., D.N. and T.J.K. receive research funding from Pharmacyclics. J.S. is a current employee of Moderna Therapeutics. J.G.G. receives grant funding from Janssen, Acerta, Celgene; and received honoraria from Abbvie, AZ, Celgene, Kite, Janssen, Pharmacyclics, Roche and Novartis. K.R.R. is on Medical Advisory Boards of Pharmacyclics, Roche/Genentech and Cellectis. J.R.B. is a consultant for Abbvie, Acerta, Beigene, Genentech/Roche, Gilead, Juno/Celgene, Kite, Loxo, Novartis, Pfizer, Pharmacyclics, Sunesis, TG Therapeutics and Verastem; received honoraria from Janssen and Teva; received research funding from Gilead, Loxo, Sun and Verastem; and served on data safety monitoring committees for Morphosys and Invectys. The other authors declare no potential conflicts of interest.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1252-x>.

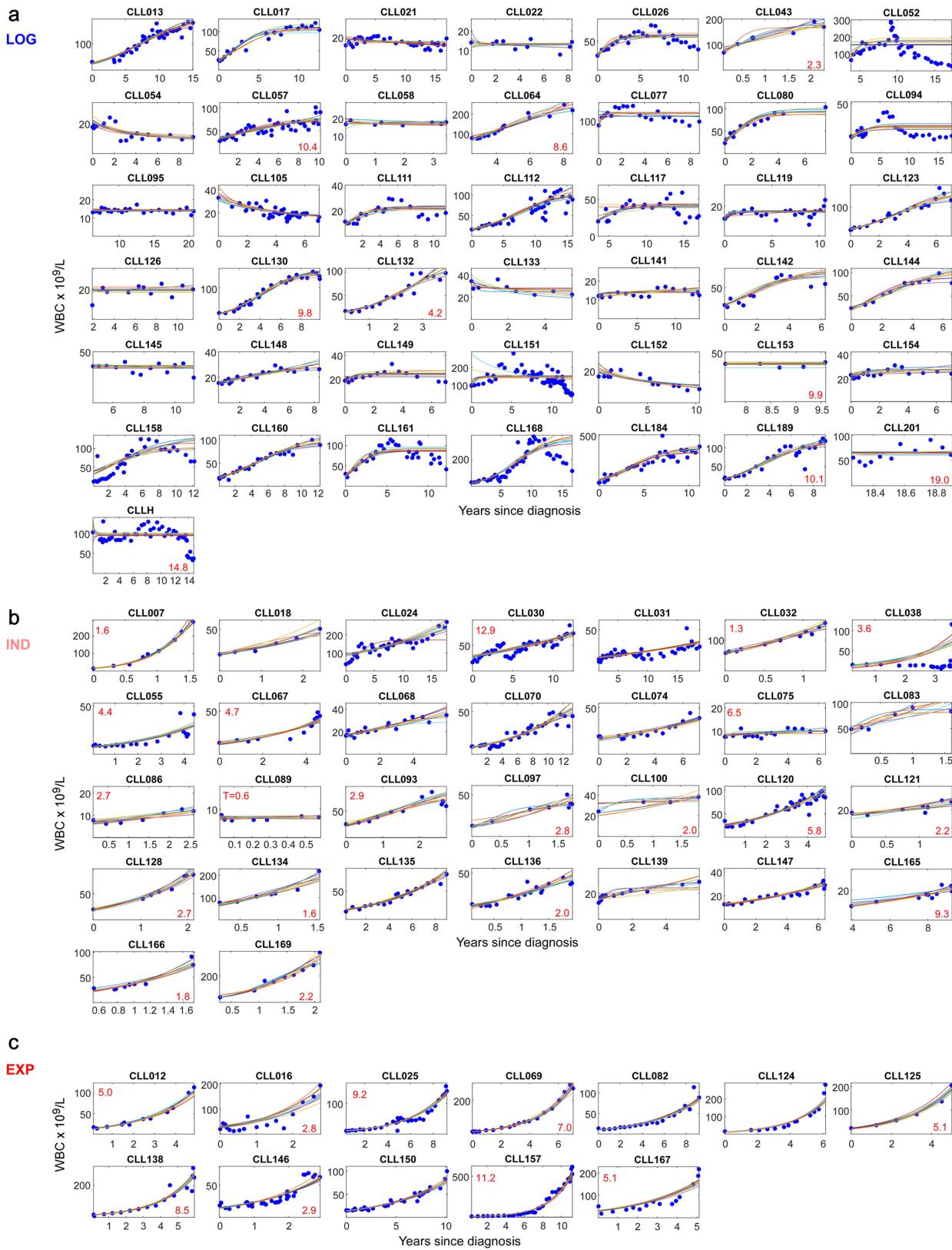
Correspondence and requests for materials should be addressed to G.G. or C.J.W.

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



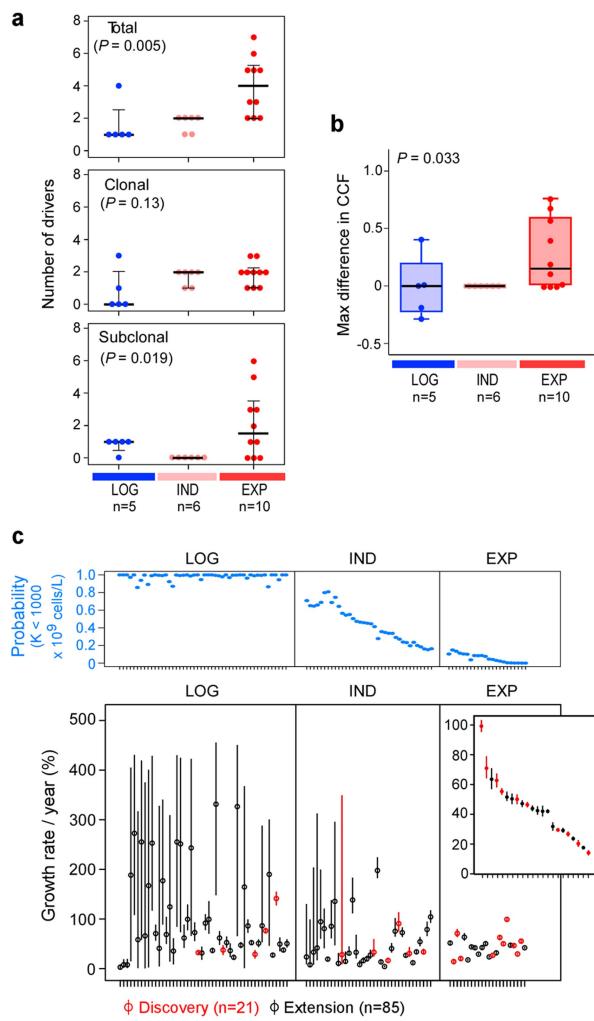
Extended Data Fig. 1 | Growth kinetics of naturally progressing CLLs from the discovery cohort. **a**, Time courses of the discovery CLL cohort (Supplementary Table 1). Circles indicate time points of samples analysed by WES. Dotted lines represent course of CLL from diagnosis (left vertical line) until last follow-up (arrows) or death (right vertical line), and solid lines indicate timeframe covered by the analysis of serial samples by WES, coloured by growth pattern. **b**, Cumulative distribution function (CDF) of posterior probabilities for carrying capacity K obtained from the Bayesian model based on a logistic growth pattern for patients. Categorizations of

the growth pattern of the individual patients are marked. **c**, Classification of patients based on the probability that their carrying capacity K is below $1,000 \times 10^9$ cells per litre (red numbers in top left corner). Also shown are the posterior probability distributions for all model parameters (carrying capacity K , growth rate r , WBC count at diagnosis X_0 and variance of the noise σ^2). Far right panels per patient: leukaemia burden information provided by WBC measurements (blue dots), with ten random fits from the Bayesian model. Red numbers in top left corners indicate time (years) from diagnosis to first treatment.

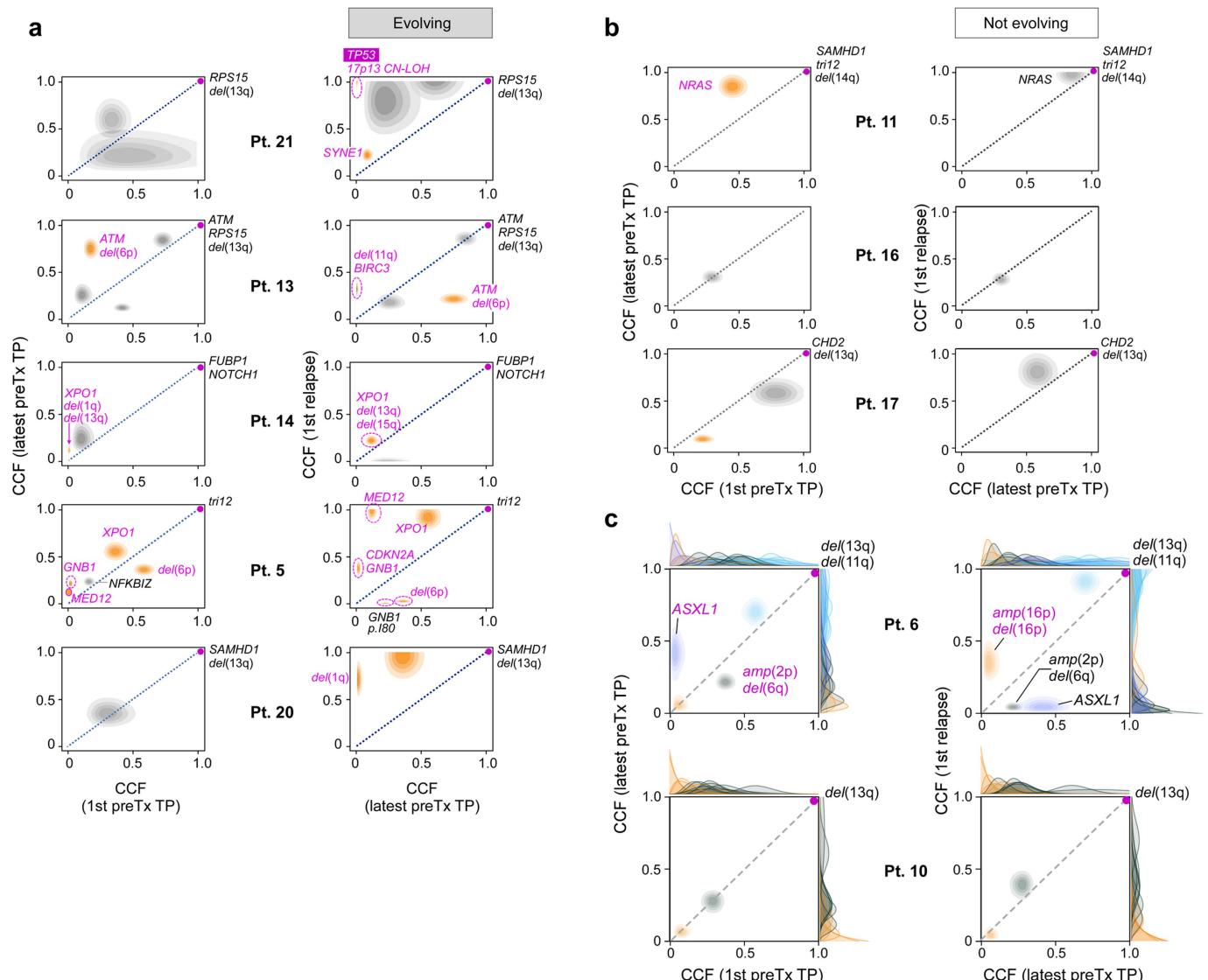


Extended Data Fig. 2 | Growth kinetics of CLLs from the extension cohort. **a–c**, Shown are samples displaying: logistic growth ($n = 43$) (a), indeterminate growth ($n = 30$) (b) or exponential growth ($n = 12$) (c). See Supplementary Table 8 for information on growth pattern fitting. Blue dots

denote WBC measurements; coloured lines denote ten random growth model fits (see Supplementary Methods). Red numbers indicate years from diagnosis to first treatment for patients who progressed to treatment.

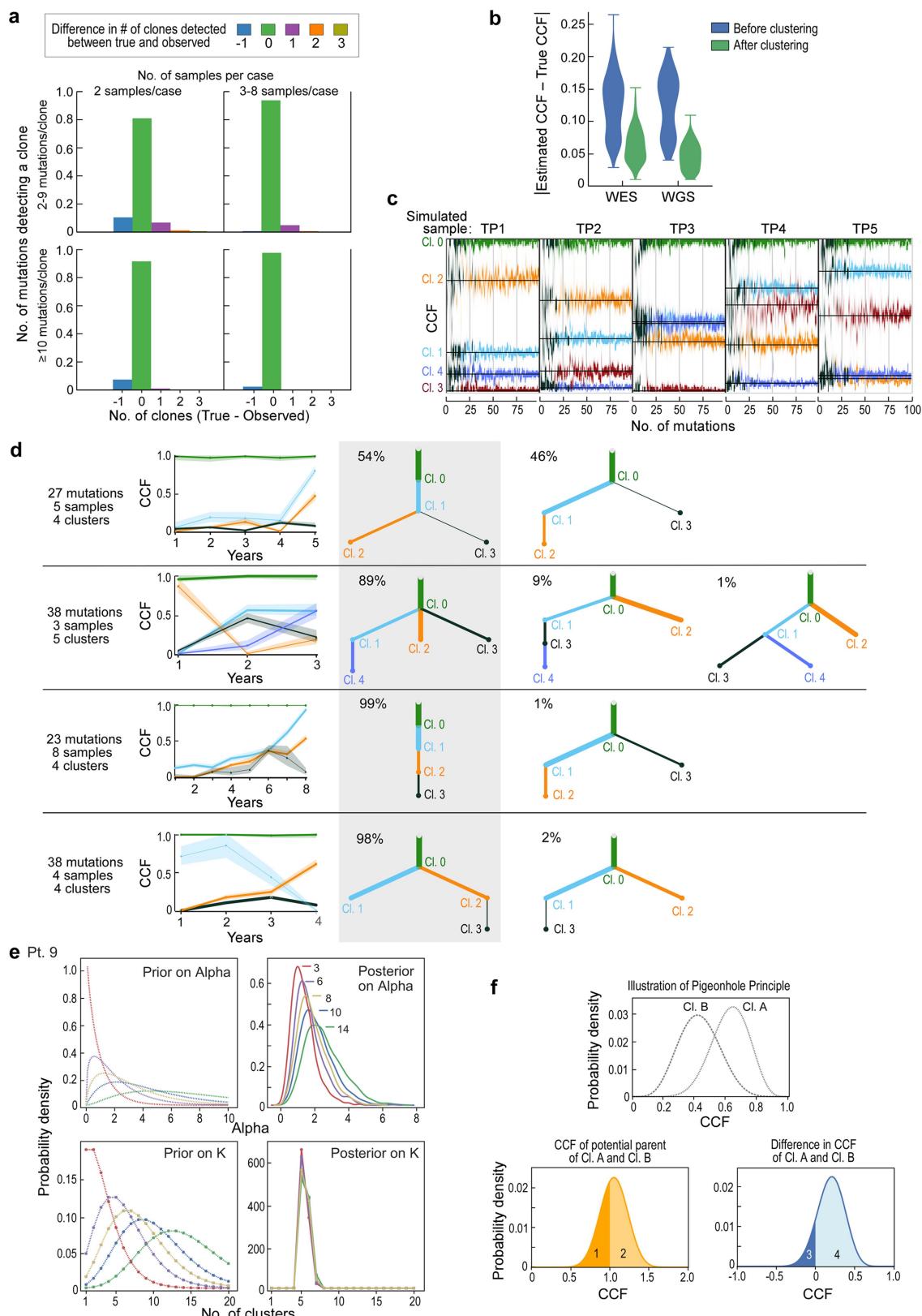


Extended Data Fig. 3 | Clonal shifts and growth rates in untreated CLL patients. **a**, The increase in the numbers of total (top), clonal (middle) and subclonal (bottom) drivers is associated with overall leukaemia growth patterns. P values were determined by Kruskal–Wallis test. **b**, A trend towards increased maximal change in the CCF of a driver event is observed between the first and last pre-treatment samples of a given patient based on growth pattern. P values were determined by Kruskal–Wallis test. **c**, Top, probability of having a carrying capacity K of WBC of less than $1,000 \times 10^9$ cells per litre (blue dots) for patients with logistic, indeterminate or exponential growth patterns. Bottom, growth rates (small circles) together with 70% credible intervals (lines) across the discovery and extension samples, ordered based on the probability of logistic growth with samples classified as displaying logistic, indeterminate or exponential growth.



Extended Data Fig. 4 | Assessment of evolutionary dynamics using sample pairs. Changes in the CCF of subclones represented as two-dimensional pair-wise plots of multi-sample clustering results. Samples at a time point (TP) closest to diagnosis (first) versus the last sample before treatment (preTx) are shown in the left column; samples at the last time point before and the first time point after treatment are shown in the right column. **a**, **b**, Patients are grouped based on those having: subclones with significant evolution (**a**) or subclones that maintain interclonal balance (**b**). Significantly evolving subclones are indicated in orange (Supplementary Methods); expanding CLL driver mutations are coloured magenta. **c**, Examples of genetic evolution from the first to last

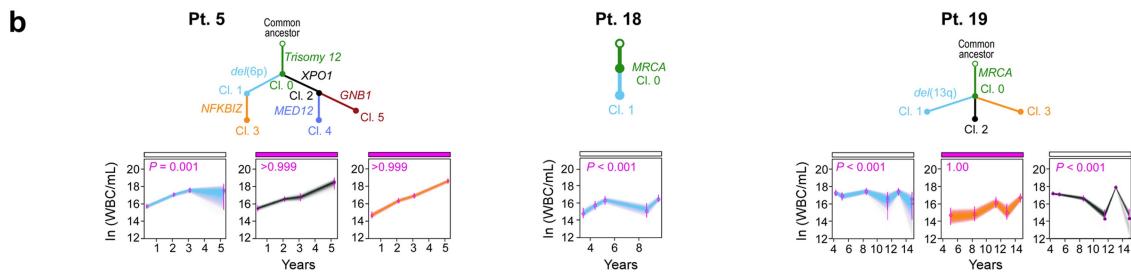
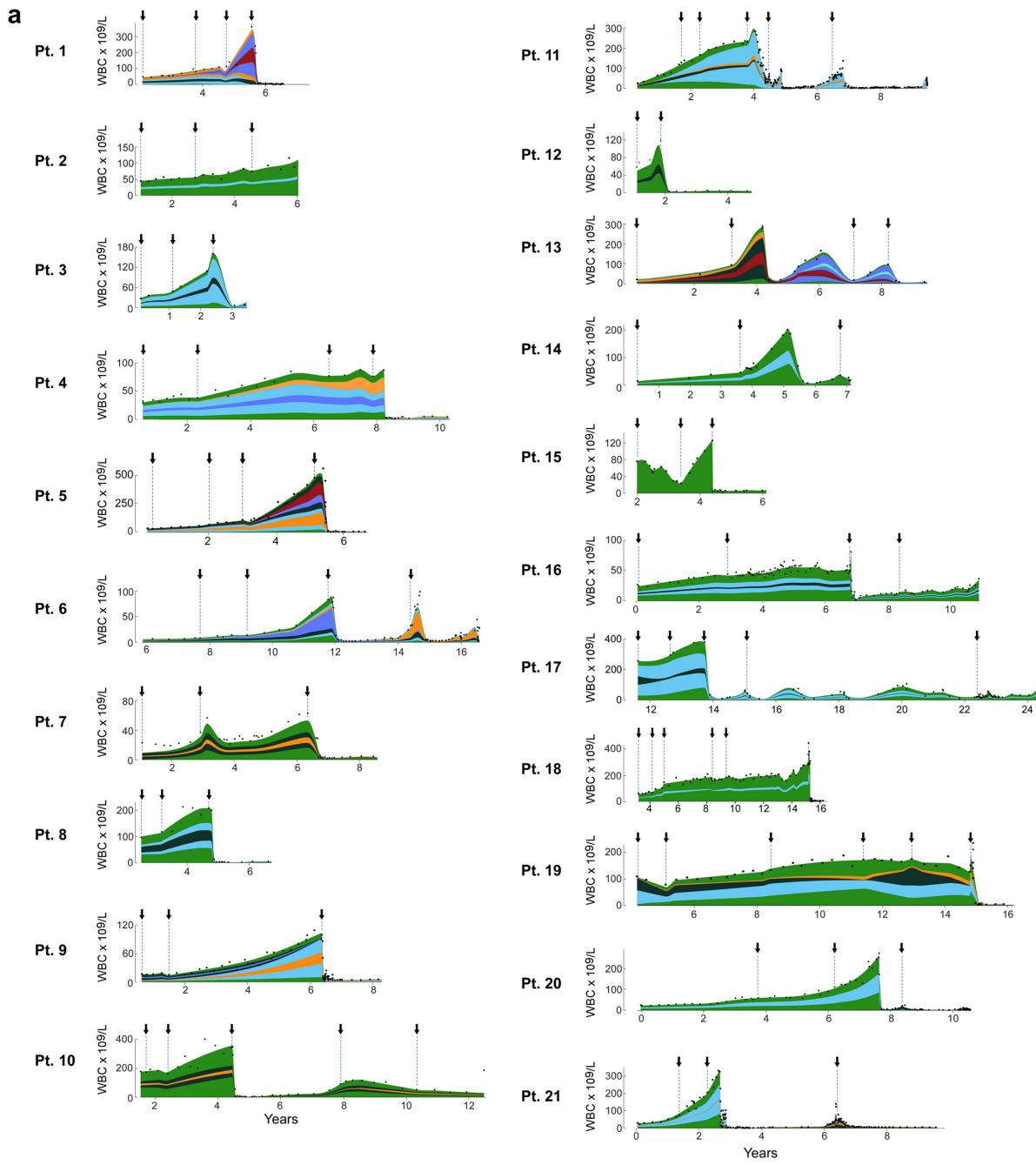
pre-treatment time points, and from pre-treatment to relapse samples for patient 6 (with significant evolution) and patient 10 (not evolving). Shown are the two-dimensional distributions that reflect the average of the positional distributions of the cluster centres along the MCMC iterations, rather than the final posterior for the cluster centre, which is determined by the normalized product of the pre-clustered distributions of the mutations that were finally assigned to each cluster. Marginal distributions (on the x and y axes) depict the CCF distributions before clustering for each individual mutation. Final cluster assignment is indicated by the colour.



Extended Data Fig. 5 | See next page for caption.

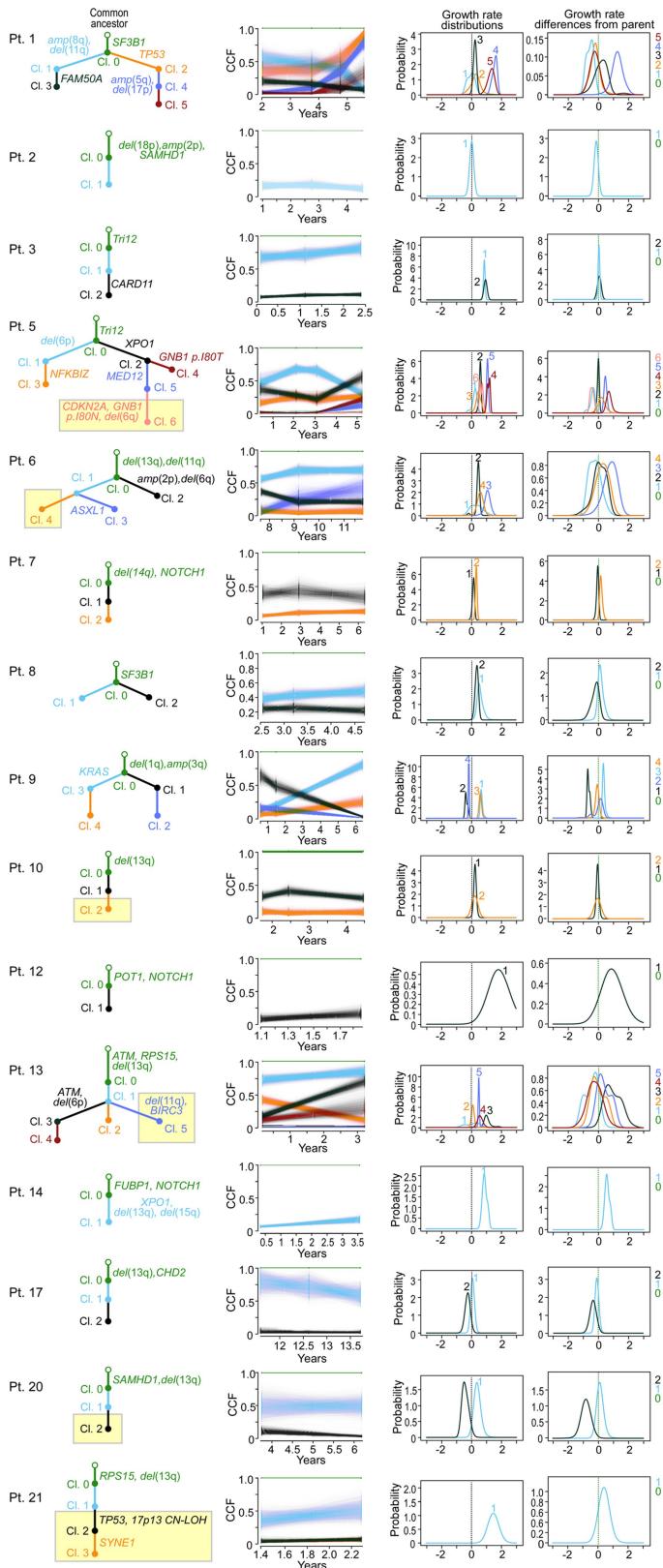
Extended Data Fig. 5 | Detecting subclones and construction of evolutionary phylogenies using simulated data. **a**, Bar plots showing the fraction of clustering results on simulated samples that are concordant with the ground truth (or differ by Δn clusters). Simulations are grouped by low (2) and high (3–8) numbers of samples per case as well as low (2–9) and high (≥ 10) numbers of mutation per subclone. **b**, Similar CCF accuracy after clustering between simulated WES and WGS data. **c**, Simulation of a case with 5 samples and 5 subclones present at different CCF levels per sample (black lines denote ground truth). The predicted CCF distributions for each cluster are plotted as a function of the number of mutations in the subclone (from 2 to 100). When the number of mutations exceeds approximately 15–20, the CCF predictions

become stable and accurate (low bias and variance). **d**, Examples of PhylogicNDT BuildTree algorithm results applied to simulated data. Grey shading highlights the correct tree, with percentage of MCMC iterations supporting the trees indicated. **e**, Analysis of prior selection for clustering. For a range of priors with varying mean number of clusters, K , the prior for α is computed, and the Dirichlet process posteriors for α and K illustrate how the choice of prior affects the estimation of K . **f**, Pigeon-hole principle: for two clusters, A and B (top), the convolution (middle) and difference (bottom) is illustrated. The area above 1.0 CCF of the convolution is consistent with the probability that they are parent–child rather than siblings. The area below 0.0 CCF of the difference represents the probability that cluster B is more prevalent than cluster A.

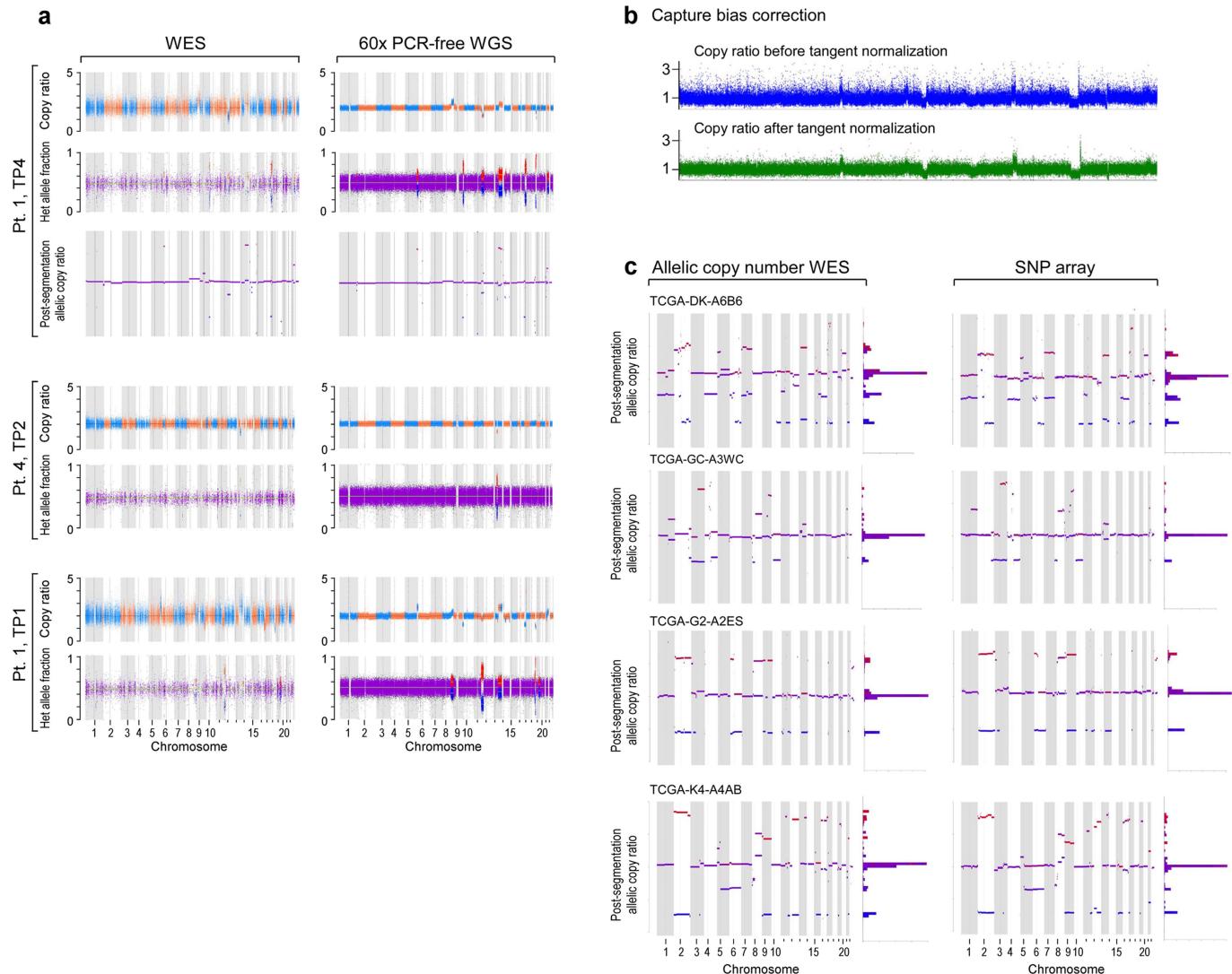


Extended Data Fig. 6 | Subclonal genetic evolutionary dynamics in the discovery cohort. **a**, Subclonal dynamics for each patient in the discovery cohort in relation to tumour load over time in the observed disease course (represented by WBC, with dots indicating an available WBC measurement). Arrows denote time of sampling with WES.

Distinguishable subclones meeting the criteria for confident detection ($>10\%$ CCF, in at least one sequenced sample) are coloured. CCFs in time periods between sequenced time points were inferred from the closest sequenced sample. **b**, Subclonal growth patterns of additional patients analogous to Fig. 4.



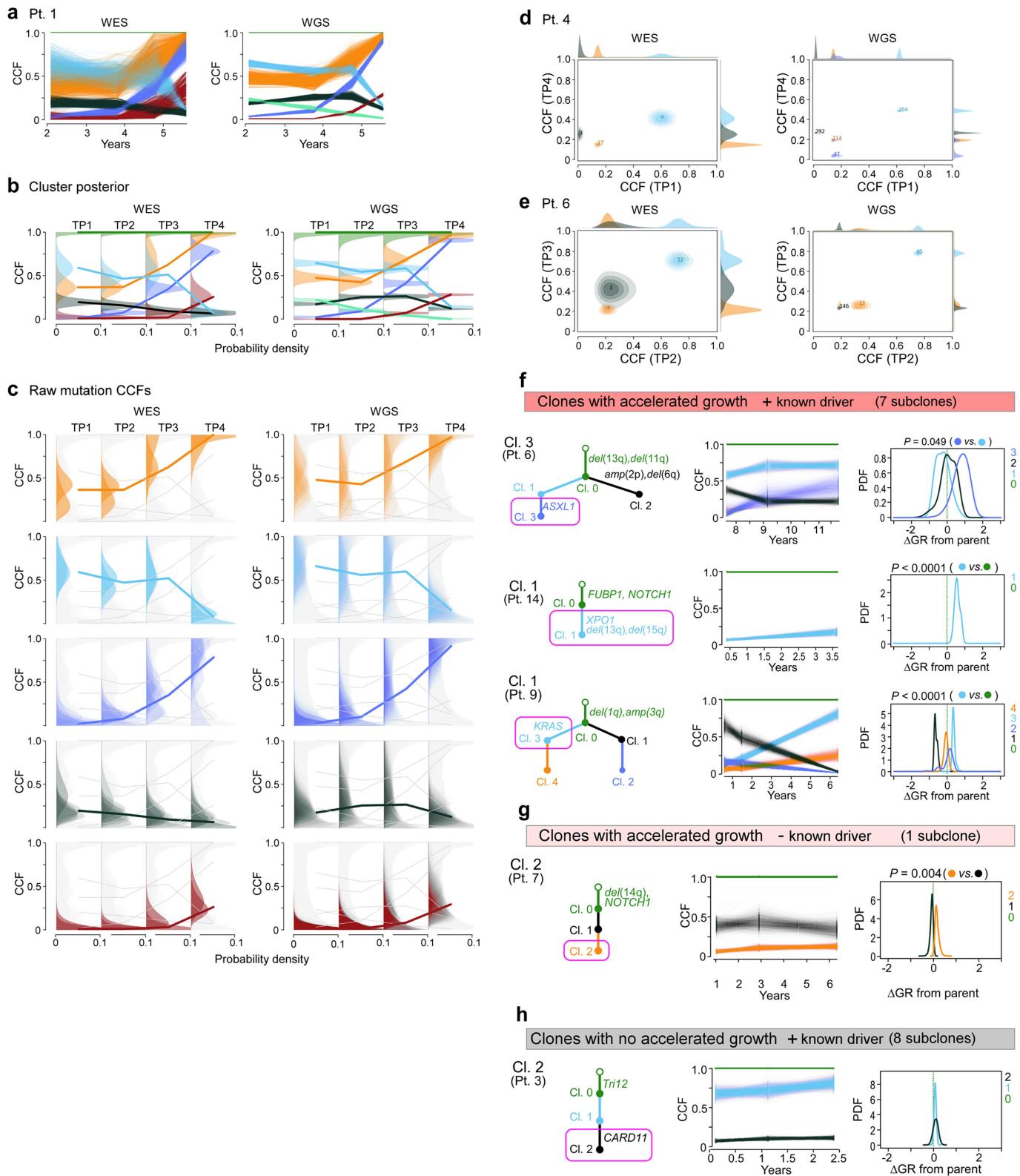
Extended Data Fig. 7 | Subclonal growth rate estimates of patients with non-bounded growth. For 15 patients with non-bounded growth (EXP and IND) and at least one macroscopic subclone, we show the following: first column: selected complete phylogenetic trees of subclones; yellow boxes indicate branches that were detectable only in relapse samples; second column: cluster CCF dynamics over time with 95% credible intervals based on uncertainty of mutation assignment; third column: pre-treatment growth rates for each generated clone within the most likely phylogeny; fourth column: relative pre-treatment growth rates of subclones compared to their respective parent subclone.



Extended Data Fig. 8 | Somatic copy number alteration calling from WES, WGS and SNP array data showing highly concordant results.
a, WES and WGS of CLLs from patients 1 and 4. **b**, Patient 1 data before

and after capture bias correction via tangent normalization³². **c**, TCGA samples with available paired WES and single nucleotide polymorphism (SNP) array data.

32. Tabak, B. et al. The Tangent copy-number inference pipeline for cancer genome analyses. Preprint at <https://www.biorxiv.org/content/10.1101/566505v1> (2019).



Extended Data Fig. 9 | Comparison of PhylogenticNDT clustering results between WES and WGS data and growth of selected subclones.

a, In patient 1, paired results of WES and WGS data were available for all four time points and demonstrate matching CCFs throughout.
 b, c, CCF posterior distributions for the cluster centres (b) and individual mutations (c) for the corresponding subclones found in WES and WGS

data of patient 1. d, e, For patients 4 and 6, two-dimensional comparisons are illustrated. f, g, Examples for subclones (magenta boxes) with a significant growth advantage relative to their parent and known driver (f), one subclone with significantly accelerated growth but no driver (g), and subclones with driver and no growth acceleration (h).

Extended Data Table 1 | Exact logistic regression modelling of the probability of treatment

	Univariate OR [95% CI]	P-value	Final multivariable OR [95% CI]	P-value*
N	79		79	
Sex, male versus female	0.94 [0.35, 2.49]	1.00		
Age ≥ 60 years vs. <60	0.34 [0.10, 1.09]	0.073		
EXP vs. LOG	12.89 [2.09, 145.73]	0.003	10.76 [1.18, 154.66]	0.032
IND vs. LOG	6.34 [2.01, 21.88]	<0.001	8.23 [1.91, 49.37]	0.002
<i>IGHV</i> unmutated vs. mutated	4.74 [1.38, 19.17]	0.010		
Del(13q) vs. no driver	0.12 [0.03, 0.41]	<0.001	0.07 [0.01, 0.33]	<0.001
Del(17p) vs. no driver	1.21 [0.15, 9.64]	0.99		
Trisomy 12 vs. no driver	3.75 [0.81, 23.84]	0.10		
Number of total drivers				
1 vs. 0	0.50 [0.009, 5.45]	0.95		
2 vs. 0	2.06 [0.44, 9.67]	0.45		
≥ 3 vs. 0	5.26 [1.59, 19.05]	0.004		
Number of clonal drivers				
1 vs. 0	0.86 [0.16, 3.81]	0.99		
≥ 2 vs. 0	4.83 [1.49, 17.26]	0.006		
Subclonal driver present vs. absent	2.31 [0.86, 6.39]	0.11		
Maximum WBC	1.00 [0.99, 1.01]	0.10		
Last WBC	1.01 [1.00, 1.01]]	0.011	1.01 [1.00, 1.02]	<0.001

Models include the 79 out of 85 patients having complete data for fluorescence in situ hybridization (FISH) cytogenetics, *IGHV* mutational status, mutations, and WBC information. From the 85, 6 patients were excluded from modelling ($n = 1$ for no sequencing, 1 for missing both *IGHV* status and FISH cytogenetics, 1 for missing *IGHV* status, and 3 for missing FISH cytogenetics). Final multivariable model included growth rate pattern, del(13q) and last WBC. The probability of treatment was measured using univariable exact logistic regression and stepwise multivariable logistic regression.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
 - Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
 - Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- Clearly defined error bars
 - State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

no software was used; data were collected from patient records in clinical trials

Data analysis

Python 2.7.13, SAS 9.4, R version 3.2.0, Matlab version R2012b

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

WES data will be deposited in dbGaP under accession code phs001431.v1.p.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Due to the exploratory nature of the study and limited availability of serial samples , no sample size calculation was performed for the discovery cohort. Criteria for composition of the discovery cohort were: CLL diagnosis with progression up to treatment according to criteria defined by the International Working Group for CLL as well as a minimum of 2 available distinct timepoints before start of treatment.
Data exclusions	No data were excluded.
Replication	To corroborate findings from the discovery cohort, we used an independent validation cohort.
Randomization	not applicable
Blinding	assessment of growth patterns and growth rates was performed prior to knowing of patient outcomes, no other blinding was performed

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics human research participants were adult patients with confirmed diagnosis of CLL

Recruitment Samples were obtained from the CLL Research Consortium tissue core <https://cll.ucsd.edu>, for which the patients were recruited from clinical trials at participating centers.