ARTICLE

doi:10.1038/nature09650

Genetic variegation of clonal architecture and propagating cells in leukaemia

Kristina Anderson[1], Christoph Lutz[2], Frederik W. van Delft[1], Caroline M. Bateman[1], Yanping Guo[2], Susan M. Colman[1], Helena Kempski[3], Anthony V. Moorman[4], Ian Titley[1], John Swansbury[1], Lyndal Kearney[1], Tariq Enver[2]† & Mel Greaves[1]

Little is known of the genetic architecture of cancer at the subclonal and single-cell level or in the cells responsible for cancer clone maintenance and propagation. Here we have examined this issue in childhood acute lymphoblastic leukaemia in which the *ETV6–RUNX1* gene fusion is an early or initiating genetic lesion followed by a modest number of recurrent or 'driver' copy number alterations. By multiplexing fluorescence *in situ* hybridization probes for these mutations, up to eight genetic abnormalities can be detected in single cells, a genetic signature of subclones identified and a composite picture of subclonal architecture and putative ancestral trees assembled. Subclones in acute lymphoblastic leukaemia have variegated genetics and complex, nonlinear or branching evolutionary histories. Copy number alterations are independently and reiteratively acquired in subclones of individual patients, and in no preferential order. Clonal architecture is dynamic and is subject to change in the lead-up to a diagnosis and in relapse. Leukaemia propagating cells, assayed by serial transplantation in NOD/SCID IL2Rγ^{null} mice, are also genetically variegated, mirroring subclonal patterns, and vary in competitive regenerative capacity *in vivo*. These data have implications for cancer genomics and for the targeted therapy of cancer.

Recent genome-wide scrutiny of cancer cells has revealed extraordinary complexity, with substantial numbers of both potential 'driver' and neutral or 'passenger' mutations per case[1,2]. Informative though these screens are, they probably reflect predominant or composite genetic landscapes that obscure the existence of subclonal heterogeneity of disease[3]. Intraclonal genetic diversity is a common feature of cancer[4] and is probably, from a Darwinian, natural selection perspective, the essential substrate for clonal evolution, disease progression, relapse or metastasis. Subclonal genetic complexity might also be an important consideration for therapeutic targeting. Furthermore, if a subset of 'stem-like' cancer cells, or, as we refer to, propagating cells, are the basis of sustained clonal expansion and disease progression[5] then, in principle, they should be genetically diverse if selection and passage through evolutionary bottlenecks is to occur.

Identifying intraclonal genetic architecture requires genetic scrutiny of single cells or clonal foci, and there are limited examples of this so far[6]; nevertheless, they testify to the existence of significant heterogeneity. The genetic diversity of cancer propagating cells is, as yet, unexplored. We elected to address this issue in lymphoblastic leukaemia. The substantial advantage of this cancer, in addition to its amenability to single-cell analysis, is that it is minimally deranged or unstable, genetically, and the broad, temporal sequence of genetic events is known. For the B-cell precursor subset of childhood acute lymphoblastic leukaemia (ALL) with *ETV6–RUNX1* fusion studied here, the latter genetic lesion is predominantly a prenatal and presumed initiating event[7]. It is coupled with a modest number (3–6) of recurrent, genomic copy number alterations (CNA)[8]. These accrue as secondary and, most likely, postnatal lesions[9] in genes that, predominantly, regulate the cell cycle or B-cell differentiation[8].

Subclonal diversity of genotypes in ALL

We initially selected 60 cases of *ETV6–RUNX1*-positive ALL and in which *ETV6* was also deleted (15–85% of cells) as detected by fluorescence *in situ* hybridization (FISH). Of these, 30 were further selected (see Supplementary Table 1) that also had (by FISH) deletion of *PAX5* ($n = 15$) or *CDKN2A* (also called *p16*) ($n = 12$) or deletions of both *PAX5* and *CDKN2A* ($n = 3$) in at least 10% of cells. All 30 cases were then scrutinized using a multiplexed combination of distinctive fluorochrome-labelled bacterial artificial chromosome (BAC) probes. Two-hundred cells with the *ETV6–RUNX1* fusion signal (that is, the reference founder mutation present in all leukaemic cells) were evaluated for each case and each individual cell designed an allele status (that is, mono- or bi-allelic deletion) for *ETV6* and *PAX5* (three colour) or *ETV6* and *CDKN2A* (three colour) or *ETV6*, *PAX5* and *CDKN2A* (four colour). The use of an *ETV6–RUNX1* probe also allowed us to detect duplication of the fusion gene (in 15 out of 30 cases) or an extra copy of chromosome 21q (via *RUNX1* signal copy number; in 21 out of 30 cases). The latter is a common genetic abnormality in ALL and an assumed driver event[10]. Cutoff levels (%) for scoring genetically distinctive subclones were determined using normal blood controls and varied depending upon probe set combination. A threshold was set at 2% for cells with a single CNA (in addition to *ETV6–RUNX1* fusion) and 1% for cells with two or more CNA (see Methods and Supplementary Table 2).

Enumeration of CNA in individual cells in reference to *ETV6–RUNX1* fusion—the universal marker of all the leukaemic cells—allowed us to identify distinctive genetic signatures of subclones and their relative frequencies. From this we could infer the most likely evolutionary or ancestral relationships between the subclones and derive a clonal architecture.

The genetic architectures that were observed were very diverse. The simplest of these genetic architectures that we identified (in 6 cases) involved two or three subclones that could be aligned in a linear sequence (Fig. 1a); however, these were cases with the lowest complement of CNA (only two or three out of the possible total of the seven pre-selected) in addition to the *ETV6–RUNX1* fusion. All

[1]Section of Haemato-Oncology, The Institute of Cancer Research, Sutton SM2 5NG, UK. [2]MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DS, UK. [3]Paediatric Malignancy Unit, Great Ormond Street Hospital & UCL Institute of Child Health, London WC1N 3JH, UK. [4]Leukaemia Research Cytogenetics Group, Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne NE1 4LP, UK. †Present address: University College London Cancer Institute, London WC1E 6BT, UK.

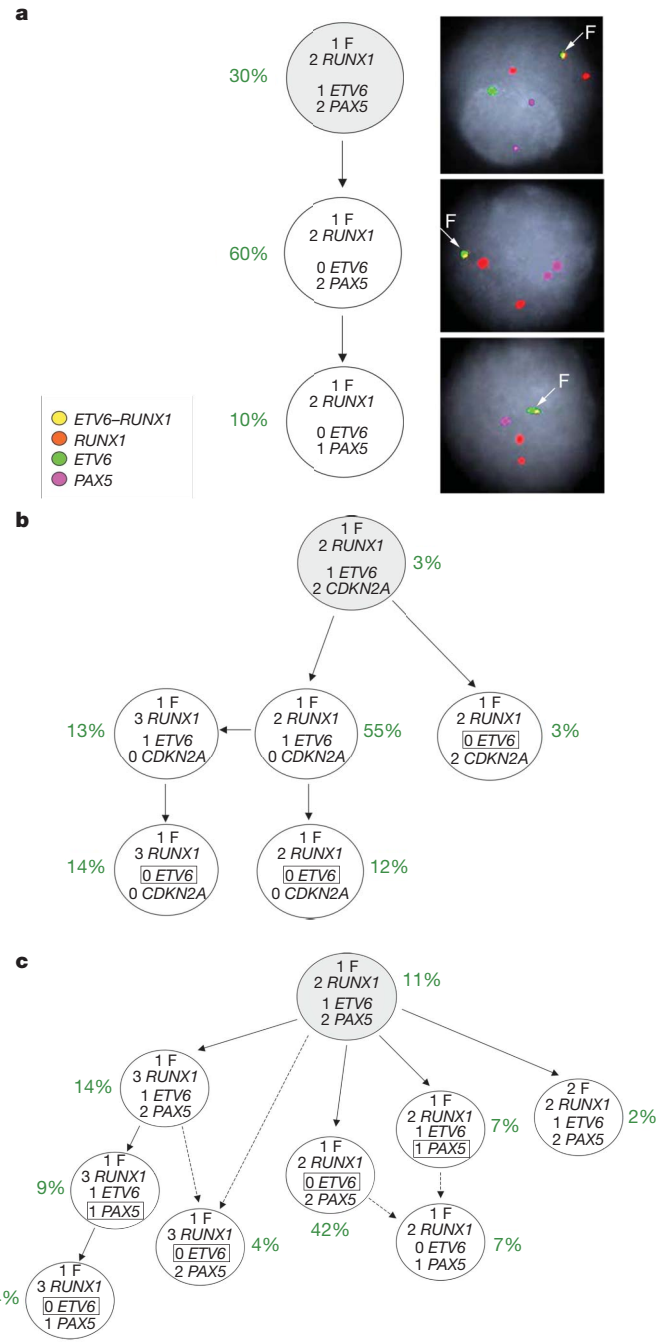

Figure 1 | Examples of subclonal architecture in ALL. **a**, Apparent linear architecture, with three clones (patient no. 13). Representative FISH images on the right show examples of each subclone. **b**, Moderately complex architecture with five subclones (patient no. 8). Loss of the untranslocated *ETV6* allele occurs independently in three separate subclones (boxes). **c**, Complex architecture with eight subclones (patient no. 16). *PAX5* deletions occur independently in two separate subclones (boxes). Arrows indicate probable (or most likely) ancestral derivation of subclones; dashed arrows indicate possible (or alternative) origins of subclones. F, yellow signal, *ETV6–RUNX1* fusion gene; 2 *RUNX1*, two red signals (one large, one small) corresponding to one normal *RUNX1* allele, and one small remnant generated from disruption of *RUNX1* allele involved in the gene fusion; *ETV6*, green signal corresponding to the normal (untranslocated) *ETV6* allele; *PAX5* (or *CDKN2A*), pink signal.

other 24 cases had a more marked subclonal heterogeneity with up to ten subclones related via a branching ancestral tree (Fig. 1b, c). Figure 1a–c illustrates examples of the clonal architectures observed (all other cases are depicted in Supplementary Fig. 2).

Inspection of clonal genotypes reveals some previously unrecognized features. It is apparent that the common or highly recurrent CNA are not acquired in any preferential order, indicating that their potency as oncogenic mutations may not be contingent upon (or epistatic to) other CNA. Subclones with the highest number of CNA, positioned 'terminally' in the branching architecture, were not necessarily numerically dominant (for example, all three cases illustrated in Fig. 1). Unexpectedly, CNA involving the same gene could be simultaneously present in distinct subclones and must therefore arise more than once, independently. *ETV6* was independently deleted two to three times in 14 of the 30 cases (see Fig. 1b, c), *PAX5* deleted two or three times in 8 out of 18 cases (see Fig. 1c) and *CDKN2A* deleted two times in 4 of 15 cases. This raises interesting mechanistic questions and suggests that these lesions are not only selected on the basis of clonal advantage but may be targeted for DNA-level breakage. One possible mechanism is via off-target effects of RAGS or AID[11–13].

Immunophenotypes and genetic diversity

There is a spectrum of early B-lineage differentiation-linked immunophenotypic signatures in ALL[14] and evidence has been presented indicating that cells with several different antibody-defined phenotypes may have leukaemia propagating activity *in vivo*[15]. We analysed the genetic heterogeneity of cells flow sorted on the basis of their expression of CD34 (immature lineage marker) or CD20 (more mature B lineage marker). Sorted populations had similarly complex genetic architectures (Supplementary Fig. 3).

Cells with the immunophenotype $CD34^{+}CD38^{-/low}CD19^{+}$ appear, so far, to be unique to ALL[16]. We previously found this pro-B/stem population, possibly non-activated or quiescent ($CD38^{-}$), to be significantly enriched in ALL propagating cells when assayed in the NOD/SCID strain of mice[17]. When purified by cell sorting, these cells (from patient no. 7) had similar genetic complexity to the bulk leukaemic population (Supplementary Fig. 3).

Clonal architecture in ALL is dynamic

These descriptions of subclonal, genetic profiles in ALL are snapshots taken at a particular time point, that is, at diagnosis. It is likely that subclonal diversity and the relative dominance of subclones varies continuously with the development and progression of disease. ALL rarely has an identified prodromal phase but occasionally (~2%) patients with ALL have an aplastic, pre-leukaemic phase a few months before a diagnosis of leukaemia[18]. We previously described one such patient with *ETV6–RUNX1*$^{+}$ ALL[19]. The diagnostic ALL cells had *ETV6–RUNX1* fusion but no *ETV6* deletion. Single nucleotide polymorphism (SNP) array screening revealed multiple deletions including *BTG1* and 11q and gain of chromosome X. We compared the clonal genotypes of cells from this patient at these two time points, spread some 7 months apart, and observed a marked shift in clonal architecture (Fig. 2a). The subclones dominating the aplastic, pre-malignant phase were relegated to minor, intermediary subclones in the overt leukaemic phase, with dominance of progeny clones that had homozygously deleted *CDKN2A*. A second patient with a prodromal, aplastic phase some 3 months before a diagnosis of ALL also showed a shift in subclonal dominance (Supplementary Fig. 4).

Treatment and subsequent relapse in ALL reorders the spectrum of genetic abnormalities detected by single gene probing[20] or SNP arrays[21] reflecting the probable selection of distinct subclones as a basis of relapse[21]. For five of the thirty selected patients (numbers 6, 9, 28, 29 and 30), we had matched diagnosis and relapse cells available for multiplexed FISH analysis. Clonal architecture at relapse was different from that at diagnosis. The comparative genetic profiles of subclones allowed us to identify the most likely subclone giving rise to the relapse, although this attribution was not unambiguous (Fig. 2b and Supplementary Fig. 2). Relapse seems to derive from either major or minor clones at diagnosis as previously suggested[21] but with a suggestion that more than one subclone might contribute to relapse

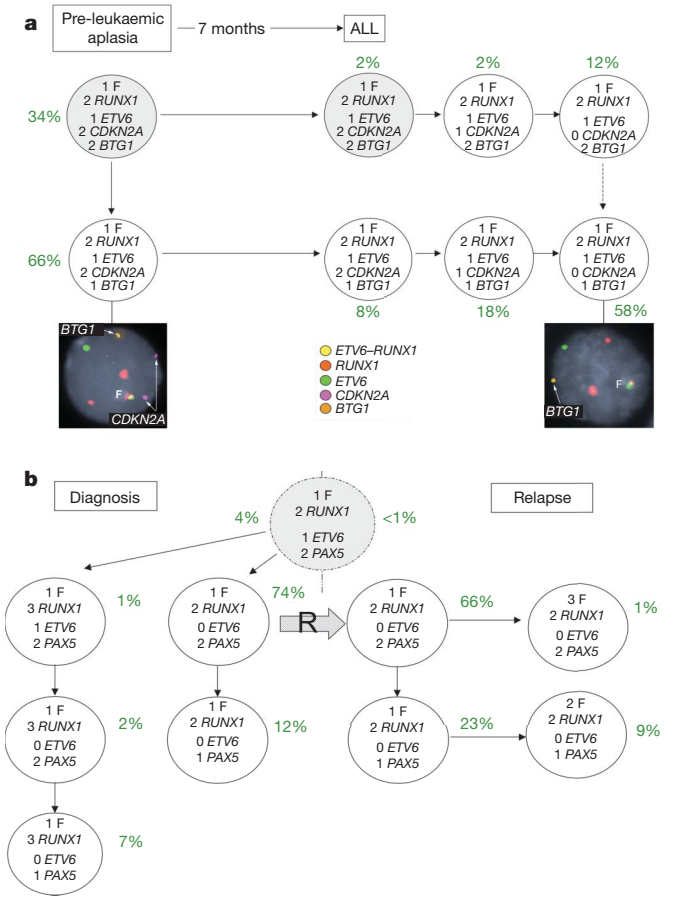

Figure 2 | Changes in clonal architecture in ALL. **a**, Scoring for *ETV6–RUNX1* fusion and simultaneous deletion of *CDKN2A* and *BTG1* in blast cells at diagnosis of ALL and in bone marrow taken 7 months earlier during a pre-leukaemic aplasia phase. During the pre-leukaemic aplasia phase no deletion of *CDKN2A* was present, although several other copy number abnormalities were present including 11q, 15q, 5q and *BTG1* gene deletions as well as gain of Xq (ref. 19). At diagnosis of ALL some 7 months later, the predominant clone contained homozygous deletion of *CDKN2A*. Only clones above the cutoffs are shown. Representative, four-colour FISH pictures for the dominant clones during the aplasia and leukaemic phases are shown at the bottom. **b**, An example of relapse originating from a major clone at diagnosis (patient no. 9) (other matched relapse cases for patients 6, 28, 29, 30 are in Supplementary Fig. 2). R, probable clonal origin of relapse.

(for example, patients 6 and 30; Supplementary Fig. 2). The data also indicate that the dominant subclone in relapse itself continues to genetically diversify, in some cases acquiring genetic lesions in the same gene (or chromosome region) as observed in primary, diagnostic subclones. This, along with previous observations on distinctive *ETV6* deletions in relapse versus diagnosis[20], provides further evidence for reiterative CNA. The patterns of genetic diversity observed in relapse indicate that genetically distinct leukaemic propagating cells can survive chemotherapy and provide a reservoir for relapse and further diversification.

Genetic diversity of propagating cells in ALL

Within the genetic architecture of ALL, it cannot be assumed that all identified subclones are self-sustaining and propagated by cells with extensive self-renewing capacity[5]. As in evolutionary speciation, it is likely that some branches or subclones are long-lived whereas others are dead ends or out-competed. Nevertheless, the architectural patterns that we observed suggested the possibility that propagating cells for ALL might also have variegated genetics and that this should be demonstrable via serial transplantation in immunodeficient mice. We transplanted, intra-tibially, varying numbers (2×10^3–10^6) of unfractionated or immunophenotypically flow-sorted leukaemic cells into pre-irradiated NOD/SCID IL2Rγ^{null} mice. Expanded leukaemic populations were re-transplanted into secondary recipient mice as a validation of self-renewal capacity. We compared the genetic signatures of the cell populations that emerged by successful regeneration *in vivo*, from first and secondary transplants, with those in the original diagnostic sample. Mice with regenerated ALL had significant proportions (3.1 to 93.5 av. 59.4; Supplementary Tables 3 and 4) of human haematopoietic (CD45$^+$) cells in the marrow and large, pale spleens (Supplementary Fig. 5b, c). Effectively, all (>99%) human CD45$^+$ cells were leukaemic with the *ETV6–RUNX1* fusion (Supplementary Fig. 5d). Genetic analysis was carried out on cells harvested from bone marrow but when assessed, spleen provided the same result (Supplementary Table 4).

Leukaemic regeneration *in vivo* was observed consistently in both unfractionated populations and in fractions defined immunophenotypically as CD34$^+$CD38$^{-/low}$CD19$^+$ and CD34$^+$CD38$^+$CD19$^+$ (Supplementary Tables 3 and 4). This accords with previous evidence[15] that propagating cells in B-cell precursor ALL are not restricted to one immunophenotypic compartment.

In all 24 mice with primary or secondary leukaemic regeneration, several genetically distinct subclones were present, the patterns of which reflected the diversity of subclones identified in the original diagnostic sample (Figs 3 and 4 and Supplementary Tables 3 and 4). The leukaemic cells regenerated in secondary transplants were compared to pre-transplant primary cells by high-resolution SNP arrays. For patient no. 3, these data confirmed subclonal loss of *CDKN2A*, subclonal gain of chromosome 21 and loss of one copy of *ETV6* in

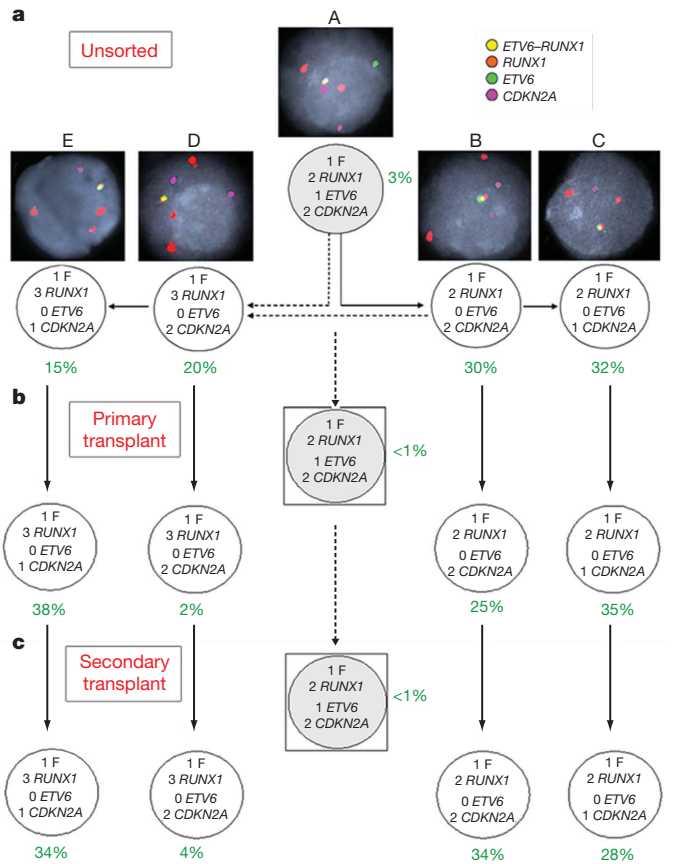

Figure 3 | Genetics of cells propagating NOD/SCID IL2Rγ^{null} mice. Leukaemic cells from patient no. 3 before injection (**a**), after primary transplantation (**b**; mouse 1, Supplementary Table 3) and after secondary transplantation (**c**; mouse 2, Supplementary Table 3). Representative FISH images are shown of the four subclones (B–E) and the putative pre-leukaemic cell (A) at the top in **a**. Boxed cell, below significance threshold.

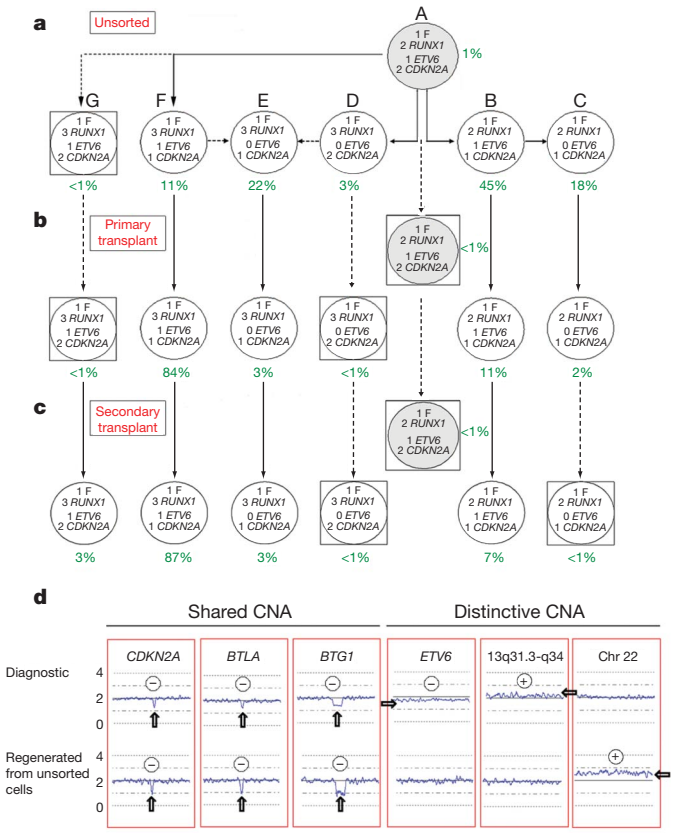

Figure 4 | Shift in clonal architecture of ALL after *in vivo* NOD/SCID IL2Rγ^{null} transplantation. Clonal architecture of unsorted cells from patient no. 7 before injection (**a**), after primary transplantation (**b**; mouse 1, Supplementary Table 4) and after secondary transplantation (**c**; mouse 2, Supplementary Table 4). **d**, Summary of SNP data on diagnostic versus secondary transplant ALL cells. The blue lines indicate the mean copy number plot of five contiguous SNPs. Middle line, normal diploid copy number. Two fractions of DNA have been examined using 500K SNP arrays: diagnostic DNA (unsorted cells from patient no. 7 before injection) and DNA from leukaemic cells regenerated in mice (mouse 2, Supplementary Table 4). Deletions of *CDKN2A*, *BTLA* and *BTG1* are present in both diagnostic and regenerated samples, whereas deletion of *ETV6*, subclonal gain of the 13q31.3-q34 region and gain of chromosome 22 are distinctive between the two samples. Plus and minus symbols indicate gains or losses of genetic regions, respectively. Arrows highlight CNA.

both diagnostic and regenerated samples (Supplementary Fig. 6 and Supplementary Table 3).

The genetic profiles of regenerated leukaemias revealed variable potency of genetically distinct subclones. For patient no. 3 (Fig. 3 and Supplementary Table 3), four subclones read-out in primary and secondary transplants with three being co-dominant. The putative pre-leukaemic clone (with *ETV6–RUNX1* fusion only) did not regenerate. With unfractionated cells from patient no. 7 (Fig. 4 and Supplementary Table 4), four of six subclones regenerated upon secondary transplantation. One of these (subclone F in Fig. 4) was dominant despite being a minor (11%) subclone in the initial diagnostic sample. SNP arrays were used to compare the primary diagnostic sample (patient no. 7) versus the regenerated (secondary) transplant leukaemias. These revealed that both leukaemic populations had *BTLA* and *BTG1* deletions (Fig. 4d) in addition to the CNA in *CDKN2A*, *ETV6* and chromosome 21. The regenerated leukaemia had an additional chromosome 22 that appeared to be absent from the initial, diagnostic cell population (Fig. 4d). This was further investigated by FISH using a chromosome-22-specific BAC probe. This confirmed the SNP array data with the majority of cells (with three *RUNX1* and one *ETV6* signals) in the regenerated leukaemia having an extra chromosome 22 signal (Supplementary Fig. 7). No cells with an extra chromosome 22 signal were detectable in the initial diagnostic sample in accord with the SNP array data. However, because clone F was dominant in all 9 of 9 mice transplanted with the same cell population (Supplementary Table 4), we assume that a minor subclone of clone F with the extra chromosome 22 was present in the diagnostic sample but at a $<1\%$ frequency. Variable competitive potency of subclonal regeneration was also seen in mice injected with immunologically fractionated cells (Supplementary Table 4 and Supplementary Figs 7 and 8).

These data are indicative of additional genetic complexity of subclones and their propagating cells. Moreover, they indicate that distinctive genotypes are associated, functionally, with variable competitive regeneration *in vivo*.

Discussion

Cancer development at the cellular level is widely regarded as a Darwinian evolutionary process involving 'natural selection' of genetically variant cells in the context of a complex micro-environmental ecology[22–24]. Mutational and phenotypic diversity between cells is, in principle, fundamental to this process. Moreover, driver mutations can be expected to have maximal selective currency when present in cells with self-renewing functionality.

Evidence for intraclonal genetic diversity in cancer has been provided by chromosome karyotype[25], by genetic analysis of multi-focal (but monoclonal) cancers[26], by FISH-based screening of tissue sections[27–29] or immuno-selected cells[30], by the molecular probing of multiple small biopsies[31] or of micro-dissected tissue[32–34] and, recently, by sector-ploidy profiling[35]. Small numbers of individual circulating tumour cells have also been scrutinized for their divergent genetic profiles[36,37]. These studies collectively testify that contemporaneous intraclonal genetic heterogeneity is commonplace and, in some cases at least, the degree of clonal diversity is predictive of disease progression[31]. Most of these data derive from epithelial carcinomas with complex genetic profiles, coupled, in most cases, to genetic instability. In such cases the historical timing and sequence of critical or driver mutational events is effectively buried and clonal architecture could be extremely complex unless clonal dominance occurs.

A common assumption for both leukaemias and cancer in general, based on the original evolutionary model of ref. 22, is that progression of disease and predominant genetic profiles reflect sequentially dominant clones and an essentially linear dynamic. Our data (summarized in Supplementary Fig. 1) suggest dynamic patterns of subclonal development and ancestral relationships that are nonlinear with a variable branching architecture. Patterns of genetic diversity in other cancers—assessed by single cell or ploidy sorted cell comparative genomic hybridization (CGH)[35,38,39], oncogenically neutral microsatellite markers[31,40] or deep-sequenced IGH gene rearrangements[41]—also indicate nonlinear, branching clonal trajectories. Collectively, these data indicate that cancer has a cellular and genetic architecture reminiscent of Darwin's iconic evolutionary tree (or bush) diagram depicting speciation[42].

The extent of genetic variegation in subclones that we detect must be a significant underestimate. We screened for a limited number of pre-selected CNA, which means that other CNA plus any sequence-based driver mutations present will not have been registered. Moreover, antecedent or intermediary subclones, below the 1–2% frequency which we set as a threshold, were identified with more extensive screening in several cases (see Supplementary Fig. 2 patients 2, 5, 15, 18, 20, 6). Identifying the full complexity of subclonal architecture and genetic diversity in ALL (and other cancers) will ultimately require whole-genome analysis at the single cell level.

Our data provide the first direct evidence for genetic diversity of cancer propagating cells within individual patients. The consistent patterns of subclonal regeneration in mice (that is, in different mice injected with the same cellular inoculum) suggest variable capacity intrinsically associated with the distinct genotypes of propagating

cells. The competitive potency of particular subclones observed, however, may to some extent reflect selective pressures exerted by regenerative stress in a murine tissue environment. Natural clonal selection in patients might produce different outcomes.

It will be important to assess if genetic diversity of propagating cells holds true for other types of leukaemia and cancer in general. If it does, then there would be significant implications for both the cancer-stem-cell concept itself and for the therapeutic targeting of such cells. The original model of a distinct, hierarchically positioned subpopulation of cancer stem cells[5] has proved contentious in both ALL[15] and other cancers[43–48]. It has been suggested that the NOD/SCID *in vivo* readout for human cancer stem cell may, at least for some cancers, simply register dominant subclones[43,48]. Or, alternatively, that cancer stem cells exist but evolve over time[44,45]. We have previously documented that 'pre-leukaemic' and overt leukaemia propagating cells in *ETV6–RUNX1*-positive ALL, although clonally related by descent, are distinctive in IgH rearrangements and phenotype[17]. Our current data fit best with what we refer to as a 'back to Darwin' model for cancer propagating cells and resultant clonal architecture[46]. In this, cells with self-renewing properties have variegated genotypes providing the units of selection in the evolutionary diversification and progression of disease. Both sequential and concurrent genotypic variation in propagating cells occur in ALL and, we predict, are likely to do so in other cancers, providing a rich substrate for progression of disease. Although it has yet to be evaluated, it is likely that genetic diversity of cancer propagating cells will be associated with both frequency variation and diversity of functional properties, for example, differentiation status, niche occupancy, quiescence and drug or irradiation sensitivity. This may help to explain some of the inconsistencies and controversies in the cancer-stem-cell field[44, 47,48,49]. Genetic diversity in cancer varies in extent with stage of disease[28,33], probably reflecting the impact of intraclonal competition and ecological bottlenecks. Single cells might negotiate very stringent bottlenecks but the genetic profiles that we observed in relapsed ALL and as recorded in, for example, prostate cancer metastases[37,50] indicate continued diversification of propagating cells and dominant or therapy-resistant subclones.

This perspective contrasts with the unidimensional or flat (albeit very complex) genetic landscapes of cancer implied in portraits derived from whole-genome scans. This architectural distinction may be of some clinical consequence. Targeted therapy, if directed at mutant molecules, may have limited efficacy if the targets themselves are not initiating lesions but secondary mutations segregated in subclones, even when the latter appear dominant. Genetic variegation of cancer propagating cells may represent a significant roadblock to effective therapy.

METHODS SUMMARY

Archival methanol:acetic-acid-fixed cytogenetic pellets from patients with *ETV6–RUNX1* fusion-gene-positive ALL were obtained from several UK hospitals, with local ethical review committee approval (CCR 2285, Royal Marsden Hospital NHS Foundation Trust). The clinical and cytogenetic data on these patients are given in Supplementary Table 1. Interphase FISH was performed as previously described[9,19].

In each case, at least 200 nuclei were scored for the presence of the *ETV6–RUNX1* fusion gene in combination with hemizygous or homozygous deletion of *ETV6*, *RUNX1*, *PAX5*, *CDKN2A* and *BTG1* and 11q, as well as gain of *RUNX1* and duplication of the *ETV6–RUNX1* fusion gene. Controls included the scoring of residual normal cells within the diagnostic sample and scoring leukaemic cells with probes hybridizing to irrelevant oncogenes (*BCR*, *ABL*) (see main Methods and Supplementary Figs 9 and 10).

NOD/SCID IL2Rγ^{null} mice that lack any B, T and natural killer cell activity were bred and maintained under sterile conditions in accordance with Home Office regulations. Transplantation of cells was by intra-tibial injections in 7–14-week-old mice after 250 cGy irradiation. Peripheral engraftment was assessed at 9–10 weeks after transplantation and if >2% mice were killed. Further analysis included the assessment of bone marrow/spleen engraftment, FISH analysis, histological analysis and serial transplantation. For serial transplantations, recovered bone marrow cells were stained with human CD45 to detect human engraftment. An equivalent of 2×10^3 to 2×10^5 human cells was transplanted by intra-tibial injections.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 May; accepted 27 October 2010.
Published online 15 December 2010.

1. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446,** 153–158 (2007).
2. Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463,** 899–905 (2010).
3. Fox, E. J., Salk, J. J. & Loeb, L. A. Cancer genome sequencing—an interim analysis. *Cancer Res.* **69,** 4948–4950 (2009).
4. Marusyk, A. & Polyak, K. Tumor heterogeneity: causes and consequences. *Biochim. Biophys. Acta* **1805,** 105–117 (2010).
5. Dick, J. E. Stem cell concepts renew cancer research. *Blood* **112,** 4793–4807 (2008).
6. Klein, C. A. *et al.* Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer. *Lancet* **360,** 683–689 (2002).
7. Greaves, M. F. & Wiemels, J. Origins of chromosome translocations in childhood leukaemia. *Nature Rev. Cancer* **3,** 639–649 (2003).
8. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446,** 758–764 (2007).
9. Bateman, C. M. *et al.* Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia. *Blood* **115,** 3553–3558 (2010).
10. Loncarevic, I. F. *et al.* Trisomy 21 is a recurrent secondary aberration in childhood acute lymphoblastic leukemia with *TEL/AML1* fusion. *Genes Chromosom. Cancer* **24,** 272–277 (1999).
11. Kitagawa, Y. *et al.* Prevalent involvement of illegitimate V(D)J recombination in chromosome 9p21 deletions in lymphoid leukemia. *J. Biol. Chem.* **277,** 46289–46297 (2002).
12. Mullighan, C. G. *et al. BCR-ABL1* lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* **453,** 110–114 (2008).
13. Feldhahn, N. *et al.* Activation-induced cytidine deaminase acts as a mutator in *BCR-ABL1*-transformed acute lymphoblastic leukemia cells. *J. Exp. Med.* **204,** 1157–1166 (2007).
14. van Dongen, J. J. M., Szczepanski, T. & Adriaansen, H. J. in *Leukemia* (eds Henderson, E. S., Lister, T. A. & Greaves, M. F.) 85–129 (Saunders, 2002).
15. le Viseur, C. *et al.* In childhood acute lymphoblastic leukemia, blasts at different stages of immunophenotypic maturation have stem cell properties. *Cancer Cell* **14,** 47–58 (2008).
16. Castor, A. *et al.* Distinct patterns of hematopoietic stem cell involvement in acute lymphoblastic leukemia. *Nature Med.* **11,** 630–637 (2005).
17. Hong, D. *et al.* Initiating and cancer-propagating cells in *TEL-AML1*-associated childhood leukemia. *Science* **319,** 336–339 (2008).
18. Breatnach, F., Chessells, J. M. & Greaves, M. F. The aplastic presentation of childhood leukaemia: a feature of common-ALL. *Br. J. Haematol.* **49,** 387–393 (1981).
19. Horsley, S. W. *et al.* Genetic lesions in a preleukemic aplasia phase in a child with acute lymphoblastic leukemia. *Genes Chromosom. Cancer* **47,** 333–340 (2008).
20. Zuna, J. *et al. TEL* deletion analysis supports a novel view of relapse in childhood acute lymphoblastic leukemia. *Clin. Cancer Res.* **10,** 5355–5360 (2004).
21. Mullighan, C. G. *et al.* Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* **322,** 1377–1380 (2008).
22. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194,** 23–28 (1976).
23. Gatenby, R. A. & Vincent, T. L. An evolutionary model of carcinogenesis. *Cancer Res.* **63,** 6212–6220 (2003).
24. Merlo, L. M. F., Pepper, J. W., Reid, B. J. & Maley, C. C. Cancer as an evolutionary and ecological process. *Nature Rev. Cancer* **6,** 924–935 (2006).
25. Teixeira, M. R. *et al.* Karyotypic comparisons of multiple tumorous and macroscopically normal surrounding tissue samples from patients with breast cancer. *Cancer Res.* **56,** 855–859 (1996).
26. Takahashi, T. *et al.* Clonal and chronological genetic analysis of multifocal cancers of the bladder and upper urinary tract. *Cancer Res.* **58,** 5835–5841 (1998).
27. Cottu, P. H. *et al.* Intratumoral heterogeneity of HER2/neu expression and its consequences for the management of advanced breast cancer. *Ann. Oncol.* **19,** 596–597 (2008).
28. Park, S. Y. *et al.* Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype. *J. Clin. Invest.* **120,** 636–644 (2010).
29. Clark, J. *et al.* Complex patterns of *ETS* gene alteration arise during cancer development in the human prostate. *Oncogene* **27,** 1993–2003 (2008).
30. Shipitsin, M. *et al.* Molecular definition of breast tumor heterogeneity. *Cancer Cell* **11,** 259–273 (2007).
31. Maley, C. C. *et al.* Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nature Genet.* **38,** 468–473 (2006).
32. Aubele, M. *et al.* Intratumoral heterogeneity in breast carcinoma revealed by laser-microdissection and comparative genomic hybridization. *Cancer Genet. Cytogenet.* **110,** 94–102 (1999).
33. Boland, C. R. *et al.* Microallelotyping defines the sequence and tempo of allelic losses at tumour suppressor gene loci during colorectal cancer progression. *Nature Med.* **1,** 902–909 (1995).

34. Geyer, F. C. *et al.* Molecular analysis reveals a genetic basis for the phenotypic diversity of metaplastic breast carcinomas. *J. Pathol.* **220,** 562–573 (2010).
35. Navin, N. *et al.* Inferring tumor progression from genomic heterogeneity. *Genome Res.* **20,** 68–80 (2010).
36. Stoecklein, N. H. *et al.* Direct genetic analysis of single disseminated cancer cells for prediction of outcome and therapy selection in esophageal cancer. *Cancer Cell* **13,** 441–453 (2008).
37. Attard, G. *et al.* Characterization of *ERG, AR* and *PTEN* gene status in circulating tumor cells from patients with castration-resistant prostate cancer. *Cancer Res.* **69,** 2912–2918 (2009).
38. Klein, C. A. & Stoecklein, N. H. Lessons from an aggressive cancer: evolutionary dynamics in esophageal carcinoma. *Cancer Res.* **69,** 5285–5288 (2009).
39. Kuukasjärvi, T. *et al.* Genetic heterogeneity and clonal evolution underlying development of asynchronous metastasis in human breast cancer. *Cancer Res.* **57,** 1597–1604 (1997).
40. Tsao, J.-L. *et al.* Colorectal adenoma and cancer divergence. Evidence of multilineage progression. *Am. J. Pathol.* **154,** 1815–1824 (1999).
41. Campbell, P. J. *et al.* Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc. Natl Acad. Sci. USA* **105,** 13081–13086 (2008).
42. Barrett, P. H. *et al.* (eds.) *Charles Darwin's Notebooks*, 1836–1844 (Cambridge Univ. Press, 1987).
43. Adams, J. M. & Strasser, A. Is tumor growth sustained by rare cancer stem cells or dominant clones? *Cancer Res.* **68,** 4018–4021 (2008).
44. Visvader, J. E. & Lindeman, G. J. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nature Rev. Cancer* **8,** 755–768 (2008).
45. Rosen, J. M. & Jordan, C. T. The increasing complexity of the cancer stem cell paradigm. *Science* **324,** 1670–1673 (2009).
46. Greaves, M. Cancer stem cells: back to Darwin? *Semin. Cancer Biol.* **20,** 65–70 (2010).
47. Maenhaut, C., Dumont, J. E., Roger, P. P. & van Staveren, W. C. G. Cancer stem cells: a reality, a myth, a fuzzy concept or a misnomer? An analysis. *Carcinogenesis* **31,** 149–158 (2010).
48. Shackleton, M., Quintana, E., Fearon, E. R. & Morrison, S. J. Heterogeneity in cancer: cancer stem cells versus clonal evolution. *Cell* **138,** 822–829 (2009).
49. Polyak, K. Breast cancer: origins and evolution. *J. Clin. Invest.* **117,** 3155–3163 (2007).
50. Liu, W. *et al.* Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer. *Nature Med.* **15,** 559–565 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is supported by specialist programme grants from The Kay Kendall Leukaemia Fund (M.G.) and Leukaemia & Lymphoma Research (M.G., T.E.) and a Deutsche Forschungsgemeinschaft fellowship LU 1474/1-1 (to C.L.). T.E. and C.L. acknowledge support from the Oxford BRC.

Author Contributions K.A. carried out the FISH analyses. C.L. and Y.G. conducted the *in vivo* experiments. C.M.B. analysed the SNP array and FISH analysis of the patient with aplasia and ALL. S.M.C. and I.T. performed cell immunostaining and sorting. F.W.v.D. provided SNP array data. H.K., A.V.M. and J.S. provided patient data and samples. T.E. advised on design and interpretation of *in vivo* experiments. L.K. supervised the FISH studies. L.K. and M.G. designed the overall study. M.G. wrote the paper with critical input from T.E., L.K. and other authors.

Author Information SNP array data have been deposited in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under accession code GSE24412. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.G. (mel.greaves@icr.ac.uk).

METHODS

Interphase fluorescence *in situ* hybridization (FISH). Archival methanol:acetic-acid-fixed cytogenetic pellets from patients with *ETV6–RUNX1* fusion-gene-positive ALL were obtained from several UK hospitals, with local ethical review committee approval (CCR 2285, Royal Marsden Hospital NHS Foundation Trust). Interphase FISH for the *ETV6–RUNX1* fusion gene was performed using a commercial LSI *TEL-AML1* extra signal (ES) probe (Vysis, Abbott Laboratories Ltd) according to the manufacturers' instructions. This probe set contains a 350-kb probe for the 5′ end of *ETV6* (exons 1–4) and a 500-kb probe covering the entire *RUNX1* gene. The FISH signal pattern for the *ETV6–RUNX1* fusion-gene-positive cells using the Vysis probe is two red (one large, one small *RUNX1* signals), one green (*ETV6* allele not involved in the translocation), one red/green (yellow) fusion signal corresponding to the *ETV6–RUNX1* fusion gene. Bacterial artificial chromosome (BAC) or fosmid probes for the *PAX5, CDKN2A*, *BTG1*, *TBL1XR1* genes, 11q and other regions of interest were obtained from the BACPAC Resource Centre, Children's Hospital, Oakland Research Institute (http://bacpac.chori.org). These were labelled by nick translation with biotin-16-dUTP or digoxigenin-11-dUTP (Roche) and hybridized in combination with the *ETV6–RUNX1* ES probe. FISH was performed by standard protocols[9,19] and labelled probes detected with streptavidin-Cy5 (bioytinylated probes) and (1) monoclonal anti-digoxigenin (Sigma), (2) horse anti-mouse IgG-Texas red (Vector Laboratories) and (3) goat anti-horse IgG-Texas Red (Jackson Immunochemicals) (for digoxigenin-labelled probes). Fluorescent signals were viewed using an Olympus AX2 fluorescence microscope equipped with narrow bandpass filters for DAPI, FITC, Spectrum orange, Texas red and Cy5. Images were captured and analysed using a charge-coupled device (Photometrics) and SmartCapture 3 software version 3.0.4 (Digital Scientific).

Establishing cutoff levels. In each case, at least 200 nuclei were scored for the presence of the *ETV6–RUNX1* fusion gene in combination with hemizygous or homozygous deletion of *ETV6*, *RUNX1*, *PAX5*, *CDKN2A* and *BTG1* and 11q, as well as gain of *RUNX1* and the *ETV6–RUNX1* fusion gene. Diagnostic slides from 26 cases were assessed for hybridization efficiency by scoring the residual normal (*ETV6–RUNX1* fusion negative) cells on the same slide (see Supplementary Fig. 9). The percentage of these cells with loss of a single *CDKN2A* or *PAX5* signal was 0–3% (mean = 1%). As a further control, we hybridized a subset ($n = 11$) of *ETV6–RUNX1* fusion-gene-positive cases with uninvolved oncogene probes *BCR* and *ABL* (see Supplementary Fig. 10). The percentage of cells with the expected normal signal pattern (two red, two green) was 96–99% (mean = 98%). Cutoff levels for each probe were established by three-colour FISH (test probe in combination with the *ETV6–RUNX1* ES probe) using three normal control peripheral blood slides per probe. As *ETV6* and *RUNX1* were scored on each slide, the values for these two probes were based on 12 slides in total. We used a cutoff = mean + 2× standard deviation. Cutoff levels for each probe (in combination with the *ETV6–RUNX1* fusion) were established using 3–12 normal control peripheral blood slides. The cutoff levels for three-colour FISH experiments are given in Supplementary Table 2. Cutoff levels for four-colour FISH were the same as above, except for Texas red probes. The cutoff for loss of one signal using a fourth probe detected with Texas red was 6.9%, because of spectral overlap between Texas red (used to detect digoxigenin-labelled probes) and Spectrum orange (used to label *RUNX1* in the commercial *ETV6–RUNX1* ES probe). However, in most cases we used three-colour FISH as a cross-check to infer whether clones below this cutoff were real (Supplementary Fig. 2). Only fusion-gene-positive cells that also showed the small extra red signal (generated by disruption of *RUNX1*) were used to calculate the relative frequencies of the various subclones.

Genome mapping analysis. Mapping analysis was performed using 500 ng of tumour DNA. DNA was prepared according to manufacturer's instructions using the GeneChip mapping 500K assay protocol for hybridization to GeneChip Mapping 250K Nsp and Sty arrays (Affymetrix). Briefly, genomic DNA was digested in parallel with restriction endonucleases NspI and StyI, ligated to an adaptor, and subjected to polymerase chain reaction (PCR) amplification with adaptor-specific primers. The PCR products were digested with DNaseI and labelled with a biotinylated nucleotide analogue. The labelled DNA fragments were hybridized to the microarray, stained by streptavidin-phycoerythrin conjugates, and washed using the Affymetrix Fluidics Station 450 then scanned with a GeneChip scanner 3000 7G.

Copy number and LOH analysis. SNP genotypes were obtained using Affymetrix GCOS software (version 1.4) to obtain raw feature intensity and Affymetrix GTYPE software (version 4.0) using the BRLMM algorithm to derive SNP genotypes. The samples were analysed using CNAG 3.0 (http://plaza.umin.ac.jp/genome), comparing tumour sample with unpaired control DNA to determine copy number and loss of heterozygosity (LOH) caused by imbalance[51]. The position of regions of LOH were identified using the University of California Santa Cruz (UCSC) Genome Browser, May 2004 Assembly (http://genome.ucsc.edu/cgi-bin/hgGateway).

Combined fluorescence immunophenotype and FISH. Bone marrow and spleen cells from leukaemic mice were cytospun and used for combined or triple-colour immunophenotype/FISH analysis as previously described[52]. Briefly, cells were air dried and fixed in acetone before incubating in primary biotinylated mouse anti-human CD45 (Clone F10-89-4) and detecting with Avidin-AMCA (Vector Laboratories). After antibody staining, the slides were hybridized with the Vysis *ETV6–RUNX1* ES fusion gene FISH probe as described above. Cells were viewed using a Zeiss Axioskop fluorescence microscope fitted with a dual bandpass FITC and rhodamine filter, as well as individual DAPI (for AMCA immunophenotype), FITC, rhodamine and Cy5. Images were captured using a charge-coupled device (Photometrics) and fluorescence signals merged and analysed using SmartCapture X software version 2.6.2 (Digital Scientific).

Cell separation, phenotyping and sorting. Total mononuclear cells were isolated by Ficoll gradient centrifugation and directly cryopreserved in DMSO for later use. After thawing dead cells were evaluated and excluded by FACS after staining with Hoechst 33258 (Invitrogen). For sorting CD34 and CD20 positive and negative subsets, samples were stained with either mouse anti-human CD34 (IgG_1, Dako) or CD20 (IgG_1, Southern Biotech) followed by anti-mouse IgG labelled with Pacific blue (Invitrogen). Nonspecific binding of antibody was assessed using mouse IgG_1 isotype control stained samples (Dako). Sorting was performed on a BD FACSAria with analysis by Becton Dickinson FACSDiVa software. Samples were gated on forward- and side-scatter plots for mononuclear cells and further gated in forward-scatter height versus area to exclude clumped cells. Before xeno-transplantation, some cells were stained with anti-CD19 PE (BD Pharmingen), CD34 FITC (BD Pharmingen) and CD38 APC (BD-Pharmingen). $CD34^+38^{-/low}CD19^+$ and pro-B $CD34^+CD38^+CD19^+$ cells were purified by flow cytometry (in this case, using MoFlo, Dako). Data acquisition and analysis were done with Summit (Dako) software. For multi-colour cell sorting 'fluorescence minus one' controls were used to determine positive and negative staining boundaries[53]. Human cells regenerating in mice were identified by staining with anti-CD45 PeCy7 (BD Pharmingen).

NOD/SCID mouse transplantation. NOD/SCID IL2Rγ^{null} mice that lack any B, T and natural killer cell activity were bred and maintained at the Weatherall Institute of Molecular Medicine animal facility in accordance with Home Office regulations. Animals were handled under sterile conditions. Transplantations of 2×10^3–10^6 cells were performed by intra-tibial injections in 7–14-week-old mice. Recipients received 250 cGy of total body irradiation before cell injection. Peripheral engraftment was assessed at 9–10 weeks after transplantation and if peripheral engraftment was >2% mice were killed. Further analysis included the assessment of bone marrow/spleen engraftment, FISH analysis, histological analysis and serial transplantation. For serial transplantations recovered bone marrow cells were stained with human CD45 to detect human engraftment. An equivalent of 2×10^3 to 2×10^5 human cells was transplanted by intra-tibial injections.

May–Grünwald Giemsa staining. The histological analysis of patient samples and mouse bone marrow was performed by May–Grünwald Giemsa staining of bone marrow smears, bone marrow cytospin preparations and spleen swabs. Slides were analysed on an Olympus BX60 microscope.

51. Nannya, Y. *et al.* A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res.* **65,** 6071–6079 (2005).
52. Kearney, L. & Colman, S. in *Methods in Molecular Biology. Leukemia. Methods and Protocols* Vol. 538 (ed. So, C. W. E.) 57–70 (Humana Press, 2009).
53. Maecker, H. T. & Trotter, J. Flow cytometry controls, instrument setup, and the determination of positivity. *Cytometry A* **69,** 1037–1042 (2006).