Title: Single cell analysis identifies *CRLF2* rearrangements as both early and late events in Down syndrome and non-Down syndrome acute lymphoblastic leukaemia

Authors: *Potter N[1], *Jones L[2], Blair H[2], Strehl S[3], Harrison CJ[2], Greaves M[1], Kearney L[1], Russell LJ[2]

[1] The Institute of Cancer Research, London, UK;

[2] Northern Institute for Cancer Research, Newcastle University, Newcastle-upon-Tyne, UK;

[3] CCRI, Children's Cancer Research Institute, St. Anna Kinderkrebsforschung, Vienna, Austria.

* These authors contributed equally to this work

Corresponding author: Dr. Lisa J Russell, Northern Institute for Cancer Research, Level 6, Herschel Building, Brewery Lane, Newcastle upon Tyne, NE1 7RU, UK.

Tel: +44 (0)191 2082235

Email: lisa.russell@newcastle.ac.uk

Running Title - *CRLF2* rearrangements are both early and late events

Conflict of Interest - All authors have no conflicts of interest to disclose.

Sources of support – Financial support from Kay Kendall Leukaemia Fund, Children with Cancer UK, Bloodwise, Wellcome Trust, Leuka and the North of England's Children's Cancer Research Fund.

Abstract

Deregulated expression of the type I cytokine receptor, *CRLF2*, is observed in 5-15% of precursor B-cell acute lymphoblastic leukaemia (B-ALL). We have previously reported the genomic landscape of patients with *CRLF2* rearrangements (*CRLF2*-r) using both whole genome and exome sequencing, which identified a number of potential clonal and sub-clonal genomic alterations. In this study, we aimed to assess when the *CRLF2*-r; *IGH-CRLF2* or *P2RY8-CRLF2*, arose during the evolution of both Down syndrome-ALL (DS-ALL) and non-DS-ALL. Using fluorescence *in situ* hybridisation, we were able to track up to four structural variants in single cells from 47 *CRLF2*-r B-ALL patients, which in association with our multiplex single cell analysis of a further four patients, permitted simultaneous tracking of copy number alterations, structural and single nucleotide variants within individual cells. We observed *CRLF2*-r arising as both early and late events in DS and non-DS-ALL patients. Parallel evolution of discrete clones was observed in the development of *CRLF2*-r B-ALL, either involving the *CRLF2*-r or one of the other tracked abnormalities. In depth single cell analysis identified both linear and branching evolution with early clones harbouring a multitude of abnormalities, including the *CRLF2*-r in DS-ALL patients.

Introduction

Acute lymphoblastic leukaemia (ALL) is defined by primary chromosomal abnormalities that drive disease development and progression and are strongly associated with outcome (1). However, the impact of sub-clonal architecture, including structural variants and mutations, is not as well defined. Initial insights into secondary genetic changes were gained through the study of monozygotic twins in which either one or both children developed ALL (2-5). These investigations identified an *in utero* origin of ALL. In particular, pre-leukaemic clones were found to harbour the *ETV6-RUNX1* fusion, which required additional abnormalities after birth to lead to overt leukaemia. Key features of the sub-clonal architecture of these cases suggested a Darwinian natural selection model to describe the process through which leukaemia presents, progresses and evades treatment (2, 3, 6).

Initiating genomic abnormalities have not been described in all subgroups of B-lineage ALL (B-ALL). A particular subtype, known as Ph-like*/BCR-ABL1*-like ALL, constitutes 10-15% of B-ALL (7, 8). It is characterised by; a transcriptional profile similar to BCR-ABL1 driven disease; high expression of the type I cytokine receptor, *cytokine receptor-like factor 2* (*CRLF2*); the presence of tyrosine kinase activating fusion genes and mutations of genes within the JAK/STAT and RAS signalling pathways (9). Deregulation of *CRLF2* occurs via two genomic alterations (*IGH-CRLF2*, *P2RY8-CRLF2*) resulting in its overexpression, however on their own they are insufficient to cause overt leukaemia (10-12). It is well documented that copy number alterations of genes involved in B-cell differentiation and cell cycle control (10, 13-15), including *PAX5, IKZF1* and *CDKN2A,* and mutations of kinases or receptor encoding genes (9, 16), in particular *CRLF2*, *JAK2*

and *IL7R,* are common in this subtype of B-ALL. However, data indicating whether *CRLF2*-rearrangements (*CRLF2*-r) are early and/or sub-clonal event are scarce (17, 18). Evidence of intratumoral heterogeneity has been revealed by a range of techniques, including conventional cytogenetics (19) and fluorescence in situ hybridisation (FISH) (20). Next generation sequencing technologies have exposed remarkable complexities in the genomic landscape of leukaemic blasts (21); coupling this approach with single cell analysis revealed additional multiple levels of heterogeneity that may further inform of treatment failure, resistance and subsequent relapse (22, 23).

In this respect, we have previously developed a multiplex Q-PCR method to target patient specific DNA alterations in flow-sorted single leukaemic cells using the BioMark HD microfluidics platform (23). This approach allows the simultaneous detection of structural variants (SVs), including translocations and fusion genes, copy number alterations (CNAs) and single nucleotide variants (SNVs) within a single cell that can be combined to illustrate clonal evolution within a bulk sample (20, 24, 25).

We have previously reported the genomic landscape of 11 *CRLF2*-r patients, using both whole genome (WGS) and exome sequencing (WES) (26), which identified a number of potential driver genomic alterations in each case. Both read counts and variant allele frequencies varied, suggesting that some aberrations were sub-clonally distributed, whilst others were clonal. In this study, we aimed to assess at what time point the *CRLF2*-r; *IGH-CRLF2* or *P2RY8-CRLF2*, arose in the evolution of both Down syndrome-ALL (DS-ALL) and non-DS-ALL. Using FISH, we were able to track up to four SVs in single cells from 47 *CRLF2*-r ALL patients, which in association with multiplex single cell analysis of a further four patients, has permitted simultaneous tracking of CNAs, SVs and SNVs within single cells.

In this study, we have observed *CRLF2*-r arising as both early and late events in DS and non-DS-ALL. Parallel evolution of discrete clones was observed in the development of *CRLF2*-r ALL, either involving the *CRLF2*-r or one of the other tracked abnormalities. In depth single cell analysis identified both linear and branching evolution with early clones harbouring a multitude of abnormalities, including the *CRLF2*-r in DS-ALL.

Methods

Sample Cohort

Patients with a CRLF2-r detectable by FISH or Multiplex Ligation-dependent Probe Amplification (MLPA) with available material were selected from six ALL treatment trials: UKALL97/99, n=24; UKALL2003, n=16; UKALLXI, n=2; UKALLXII, n=6; UKALLR3, n=1; ALL-BFM 2000, n=2 (Supplementary Table 1). Ages ranged from 1-54 years (median 5 years, mean 9.73 years).

Institutional review board approval was obtained at each of the collaborating centres. Informed consent was obtained in accordance with the Declaration of Helsinki.

Genomic target selection

WGS and WES data for the patients included in this study have been previously published (26). We used these data to identify SVs, which could be tracked in single cells by FISH and focal CNAs and SNVs for tracking in single cells by multiplex Q-PCR. Two DS-ALL patients (19599, 11538) and two non-DS-ALL patients (11543, 21819) were investigated by single cell multiplex Q-PCR(23). SVs tracked in each patient by FISH are shown in Supplementary Table 1 and SVs, CNAs and SNVs tracked by Q-PCR are shown in Supplementary Tables 2 and 3.

Xenograft transplants

All *in vivo* studies were performed by personnel holding a current Personal Licence under the Animals (Scientific Procedures) Act 1986 and were conducted in line with current Home Office regulations, compliant with the 3R principles (Home Office license number PPL 60/4552). No further ethical approval was required. Animals were housed under pathogen free conditions, and all experimental manipulations were performed on anaesthetized mice under sterile conditions in a laminar flow hood.

Male and female NSG mice (Jax® mice strain name: NOD.Cg-Prkdcscid Il2rgtm1wjl/SzJ) were transplanted with $1x10^6$ patient primary bone marrow cells, injected intrafemorally into the right femur, as previously described (27). Animals were maintained until they showed clinical signs (weight loss, palpable spleen), which necessitated humane killing (range 3-10 months post injection). Patient-derived xenograft cells (PDXs) from the bone marrow and spleen were harvested, passed through cell strainers (0.40µm) and frozen for long term storage in 90% foetal calf serum and 10% DMSO in liquid nitrogen.

Cytogenetics and fluorescence in situ hybridisation

Fixed cells were available for 47 *CRLF2*-r patients (46 at diagnosis and 1 at relapse): 14 patients were DS (*P2RY8-CRLF2*, n=12; *IGH-CRLF,2* n=2) and 33 were non-DS (*P2RY8-CRLF2*, n=25; *IGH-CRLF2*, n=8) (Supplementary Table 1). Where possible, direct fixed cell cultures were tested to prevent discrepancies caused by outgrowth of normal cells and death of malignant cells. Multiple colour interphase FISH was used to assess the simultaneous presence of both *CRLF2-r* (XX-87136C11, XX-82904A1, RP4-674K4, RP11-309C18)(10) and deletions of *CDKN2A/B* (RP11-149I2) and/or *PAX5* (RP11-469D03) and/or *IKZF1* (G248P800745C8/WI2-3001F15) (previously identified by MLPA and single nucleotide polymorphism (SNP) arrays (26)) in the same cell, as previously

reported (10, 28). Briefly, home-grown FISH probes were mixed 1:1 with hybridisation buffer (company) and denatured at 75°C for five minutes followed by hybridisation at 37°C overnight. Coverslips were removed in 2x SSC and slides washed in 0.02% SSC with 0.003% NP40 at 72°C for two minutes followed incubation in 0.1% SSC at room temperature for two minutes. Slides were mounted with 10ul DAPI (Vector laboratories, California, USA). Automated capture and scoring was performed using an automated Olympus BX-61 8-bay stage florescence microscope with a x60 oil objective. Images were stored and analysed using the CytoVision 7.2 SPOT counting system (Leica Microsystems, Gateshead, UK). Where possible, more than 100 nuclei were scored for each FISH test by two independent analysts. A cut-off threshold of >8% was used for all multiple colour probe combinations to allow for interference and obscuring of signals. The cut-off level was established by counting the number of abnormal (false positive) signals generated when the multiple colour probe combinations were hybridised to normal cells.

Commercial and custom primers for Q-PCR and digital PCR

We aimed to track all identified somatic SVs and SNVs detected by WGS and WES encompassing those present at high and low variant allele frequencies; some could not be tracked due to the sequence surrounding the rearrangement (Supplementary Tables 2 and 3).

Custom Taqman Q-PCR assays for SV or SNP that could distinguish the mutant allele from its wild-type counterpart were designed as previously described (23). DNA copy number Taqman assays were purchased from Applied Biosystems. Three CNA assays were chosen within each DNA target region of interest and the diploid reference region encompassing *B2M* (Supplementary Table 4).

FACS for single cell collection

Diagnostic patient samples and those previously harvested from xenograft transplants were thawed from liquid nitrogen and bulk cells were labelled with carboxyfluorescin diacetate, succinimidyl ester (CFSE) and 6 diamidino-2-phenylindole (DAPI) to identify live and dead cells, respectively. Cells retrieved from successful transplantations into NSG mice were also labelled with phycoerythrin (PE)–conjugated anti-human CD45 and allophycocyanin (APC)–conjugated anti-mouse CD45 antibodies (BD Biosciences) before resuspension in PBS and DAPI in order to identify and sort the human leukaemic cells.

Single Cell Sorting and multiplex Q-PCR analysis

Single cell sorting was carried out after staining, according to our established protocol, as previously described (23). Briefly, single cells were sorted from each case into individual wells of a 96 well plate and lysed. The DNA target region of interest, including patient specific gene fusions, SNVs and CNAs, was amplified. We collected 252-336 cells from the diagnostic ALL samples, 81-252 cells from PDX and 48 cord blood cells (normal diploid control). A cell was called positive for a SNP (or SNV) if the Q-PCR cycle threshold (C_T) value was below 28. We used a modified version of the $\Delta\Delta C_T$ method (Applied Biosystems, Life Technologies Ltd.) to integrate results from multiple Taqman assays targeting the same region to determine the relative copy number for each locus. The use of multiple assays to target one region increased the accuracy of attributed CNAs. The resulting reaction mix was then diluted and Q-PCR completed using the 96x96 dynamic array and the BioMark™ HD from Fluidigm.

Several approaches were adopted during this experiment to optimise and confirm the presence of single cells and ensure that all assays performed efficiently under

192 experimental conditions (23). Assay error rates in these experiments were zero. Single
193 cell data removed from the Q-PCR analysis included those from wells with no data (no
194 cell) and those wells in which all *B2M* assays did not have a strong signal (<25 C_T). On
195 average this accounted for the removal of ~9% of data points. Only mutational spectra
196 occurring in more than one cell were included in the analysis.

197 **Results**

198 To explore whether *IGH-CRLF2* and *P2RY8-CRLF2* were early or late events in the
199 evolution of B-ALL, we performed multiple colour FISH on 47 ALL patient samples to
200 track the *CRLF2-r (IGH-CRLF2*, n=10, *P2RY8-CRLF2*, n=37) and deletions common in
201 ALL: *IKZF1* and/or *CDKN2A/B* and/or *PAX5*.

202 *IGH-CRLF2* is an early event in B-ALL

203 In those patients with *IGH-CRLF2* bone marrow blast counts at diagnosis correlated
204 with the percentage of abnormal blasts detected by multiple colour FISH
205 (Supplementary Table 1), suggesting that the tracked abnormalities were present in
206 the major leukaemic clone. In 80% (8/10), *IGH-CRLF2* was a clonal event, either
207 presenting first in the majority of cells in four patients, or together with other tracked
208 abnormalities in four patients (examples of FISH signal patterns are shown in Figures
209 1A and 1B).

210 *P2RY8-CRLF2* is an early event in B-ALL

211 In 73% (27/37) of patients with *P2RY8-CRLF2*, the bone marrow blast counts at
212 diagnosis correlated with the percentage of abnormal blasts detected by multiple
213 colour FISH (Supplementary Table 1). *P2RY8-CRLF2* presented as the first tracked
214 abnormality in 9/27 patients and together with other abnormalities in 18/27 patients
215 (for one case only relapse material was available; 6897) (examples of FISH signal

patterns are shown in Figures 1C and 1D).

CRLF2-r are observed as sub-clonal events in B-ALL

While most patients with *IGH-CRLF2* have the rearrangement as an early event in leukaemogenesis, in two *IGH-CRLF2* patients (3789 and 3141), loss of *IKZF1* was detected as the earliest event with *IGH-CRLF2* observed only in a sub-clone (example from patient 3789 is shown in Figure 2A). Patient 3141 had two subsequent relapses; at second relapse, the *IKZF1* deletion was present in 64% of cells compared to 33% for *IGH-CRLF2* as assessed by individual FISH tests (data not shown). These data confirm the sub-clonal nature of *IGH-CRLF2* in this patient.

Similarly, the sub-clonal nature of *P2RY8-CRLF2* was observed in 24% (9/37) of the patients, where the percentage of abnormal cells detected by FISH was notably less than the blast count at diagnosis, suggesting that the earliest events remained undetected in these patients (Supplementary Table 1). Of particular interest was patient 2017, where monoallelic loss of *CDKN2A/B* and *PAX5* preceded the formation of the fusion (Figure 2B). Of the two or three abnormalities tracked by FISH, *P2RY8-CRLF2* presented first in 4/9 patients and together with other abnormalities in 5/9 patients.

Parallel evolution occurs in the development of *CRLF2*-r ALL

In five *P2RY8-CRLF2* patients, evidence of parallel evolution of cells containing abnormalities of one or more genes was observed (Figure 3A-E). In three patients (12200, 3173, 4954), four sub-clones were identified. In patients 12200 and 3173, (Figure 3A and B, respectively), clones harbouring either the *P2RY8-CRLF2* fusion or loss of *CDKN2A/B* or *IKZF1*, respectively, occurred in sub-clones. The main clone seen at diagnosis must have evolved from a cell that acquired both abnormalities giving it

a competitive advantage over the other sub-clone. In patient 4954 (Figure 3C), the *P2RY8-CRLF2* fusion was present together with either, loss of one or two copies of *CDKN2A/B* and loss of one copy of *IKZF1*. These data suggest that *CDKN2A/B* undergoes deletions in independent sub-clones. The main clone seen at diagnosis must have evolved from a cell that acquired all three abnormalities giving it a competitive advantage over the other sub-clone.

In patients 20753 and 5817, the parallel clones were defined by the *CRLF2*-r. In patient 20753 (Figure 3D), one sub-clone gained an extra *P2RY8-CRLF2* fusion with subsequent loss of a single copy of *PAX5* prior to the loss of the second copy of *IKZF1*. In patient 5817 (Figure 3E), the sub-clones harboured either a *CRLF2*-r with loss of one copy of *IKZF1* (9%) or *P2RY8-CRLF2* alone (17%). Identical parallel evolution was found in patient 21572 (Figure 3F). The parallel clones were defined by an *IGH-CRLF2* translocation alone or a *CRLF2*-r coupled with loss of one copy of *IKZF1*. Together these data suggest the sub-clonal architecture observed in these cases could only have occurred if *IKZF1* and/or *CDKN2A/B* undergo deletions in independent sub-clones.

Single cell multiplex Q-PCR identified linear and branching development of *CRLF2*-r ALL

Whilst FISH allowed the detection of a small number of large SVs, it was not possible to investigate associations between small CNAs and SNVs. Single cell multiplex Q-PCR allowed a more comprehensive analysis of four patients. Individual cells were sorted from the leukaemia sample and assayed by multiplex Q-PCR for the presence of specific genetic aberrations previously identified from WGS and WES (26). A similar assessment was carried out using unsorted PDX ALL cells from the same patients in order to determine self-renewal properties of discrete clones. From these samples

and expanded single cell data, we were able to define detailed clonal architectures in which genetically distinct sub-clones were characterised by SVs, SNVs and CNAs (Figure 4).

In non-DS-ALL patients 11543 and 21819 a linear architecture was observed (Figure 4A and B). In patient 11543 (Figure 4A) the I*GH-CRLF2* translocation was secondary to multiple CNAs and SNVs, including an initial *IKZF1* deletion, with a second *IKZF1* deletion occurring after the acquisition of the *IGH-CRLF2* translocation. The architecture was recapitulated in both xenograft bone marrow and spleen cells, suggesting that all sub-clones possessed self-renewal properties. In patient 21819 (Figure 4B) the *P2RY8-CRLF2* fusion was secondary to loss of *IKZF1* and *CDKN2A* (Figure 4C). Serial xenograft transplants were ultimately populated by the major *P2RY8-CRLF2* containing clone seen at diagnosis.

In contrast, the two DS-ALL patients, 19599 and 11538, showed a branching sub-clonal architecture where the *CRLF2*-r was one of several structural alterations that defined the earliest identified clone (Figure 4C and D). In patient 19599 (Figure 4C) the sub-clonal architecture was defined by discrete *IKZF1* and *JAK2* mutations. A single evolved clone from the diagnostic sample engrafted into both spleen and bone marrow of transplanted mice, suggesting that the other sub-clones had a reduced self-renewal ability. In patient 11538 (Figure 4D) the sub-clonal architecture was defined by multiple *IKZF1* events and additional mutations that were secondary to the *P2RY8-CRLF2* fusion. Whilst the architecture was recapitulated in the xenograft, the major clones at diagnosis were present as minor clones in the xenografts.

Whilst multiple mice were not analysed by single cell QPCR for each sample, FISH was completed and showed comparable results (data not shown).

Discussion

In several blood cell cancers there appears to be a preferential order of mutations, including in ALL (20, 21, 29, 30), AML (31) and MDS (32-34). To determine whether *IGH-CRLF2* and *P2RY8-CRLF2* were early or late events in the evolution of B-ALL, we carried out single cell analysis using both FISH and multiplex Q-PCR of 51 DS and non-DS-ALL patients. Forty-seven were investigated by FISH for specific rearrangements including *P2RY8-CRLF2* or *IGH-CRLF2* coupled with deletions of *IKZF1, PAX5* or *CDKN2A/B*. The remaining four cases underwent a more detailed analysis using multiplex Q-PCR for multiple patient specific SVs, SNVs and CNAs.

CRLF2-r were observed as common early events in the majority of patients studied, including DS and non-DS-ALL patients. However, the *P2RY8-CRLF2* fusion was also found to be sub-clonal in approximately one quarter of patients investigated. The sub-clonal nature of the *P2RY8-CRLF2* fusion has previously been reported (17). In contrast, the early nature of both *IGH-CRLF2* and *P2RY8-CRLF2* in DS-ALL was recently reported, where 92% (11/12) of patients retained these rearrangements at relapse. The authors suggested that both rearrangements were early events in leukaemogenesis which may play important roles at relapse (18). However, other evidence suggests that relapse in ALL can originate from sub-clones distributed anywhere in the phylogenetic architecture of the cancer (20, 35-37), indicating that the preservation of any individual genetic lesion in relapse does not necessarily reflect a founder , early or truncal status.

In the present study, six patients showed the potential development of two independent leukaemias with clones showing parallel evolution driven by reiterative

CNAs within the same genes. In three patients with *P2RY8-CRLF2* loss of one copy of the other tracked gene occurred in discrete populations. The bulk of the leukaemia then evolved from a cell acquiring both abnormalities. In the remaining three patients the parallel clones were defined by distinct early *CRLF2*-r. The presence of reiterative genetic changes has been reported in ALL before (20, 38) and they are thought to arise by RAG-mediated mutagenesis (22). Reiterative mutation is a relatively common feature in other cancers (39) reflecting convergent (or parallel) evolution in the context of common selective pressures. It is also important to consider that we are analysing one sample at a single time point in the development of this disease. It is not always possible to determine the precise temporal order in which events take place during tumour development unless representative populations of all ancestral clones remain at diagnosis. In those where we observed the *CRLF2*-r and deletions of other genes as the earliest identifiable events, it is possible that they arose sequentially with the earliest events no longer being present at diagnosis. The limited sensitivity of FISH for detecting these rare clones impacts how precise we can be in mapping the temporal order of tumour development.

Sub-clonal heterogeneity and clonal selection has been studied in many haematological diseases, highlighting the importance of understanding sub-clonal architecture in relation to therapeutic decisions regarding individual patients (38, 40, 41). Single cell multiplex Q-PCR in two DS-ALL patients revealed the presence of *IGH-CRLF2* or *P2RY8-CRLF2* in the earliest clone, together with a multitude of SNVs and other CNAs. These patients showed a complex branching tree structure with reiterative deletions and mutations occurring in different sub-clonal populations. In contrast, the leukaemia in the two non-DS-ALL patients appeared to evolve in a linear

non-branching manner with *IGH-CRLF2* or *P2RY8-CRLF2* occurring on a background of SNVs or CNAs and arising later in leukaemia development. These data suggest that *IGH-CRLF2* and *P2RY8-CRLF2* could be earlier events in DS-ALL compared to non-DS-ALL. Although our results suggest that there is no difference between DS and non-DS ALL, we are - despite simultaneously tracking several alterations - still likely underestimating the complexity of clonal phylogeny.

The application of single cell Q-PCR to PDX cells demonstrated that in three of the four patients, the majority of clones identified at diagnosis had leukaemia propagating capacity, being present in both bone marrow and spleen in primary and second generation mice. In the PDX cells from patient 19599, only one sub-clone engrafted, suggesting that either the other sub-clones did not have self-propagating capacity, or that they were below the level of detection within the diagnostic sample. Such findings have been previously reported, in which analysis of CNAs from mice injected with as few as 100 cells remained highly related to the diagnostic sample, with only a few novel deletions arising in the primary mice (42). In some samples, clonal outgrowth of the dominant diagnostic clone was observed; however, in others they observed outgrowth of sub-clones (42). The outgrowth of certain clones would not necessarily mean the clones that are no longer present are incapable of self-renewal, but are less suited to the murine environment.

Evolution of cancer was initially assumed to be driven by a steady accumulation of genomic abnormalities over time. However, others have suggested that the presence of explosive changes caused by global genomic instability (43), a chromothripsis event (44) or the effect of a single high impact mutation (45, 46) may be responsible. Previous work has postulated that additional copies of chromosome 21 can promote

genomic instability (47, 48). Interestingly, among the four patients studied here by multiplex Q-PCR, three showed a large number of abnormalities in the earliest clone (range 11-20), suggesting that either a single or series of explosive events may have occurred creating a backbone of aberrations upon which further evolution could take place. Notably these three patients had either constitutional or somatic gain of chromosome 21 either.

In summary, our data indicates that *CRLF2*-r co-operate with multiple additional genetic alterations in ALL and that there appears to be no major restraint on whether *CRLF2*-r arise early as a founder or truncal event or later in clonal evolution.

Acknowledgements

The authors would like to thank The Kay Kendall Leukaemia Fund, Leuka, Children with Cancer UK and Bloodwise (formerly Leukaemia and Lymphoma Research) for financial support. Lisa J Russell has a John Goldman Fellowship from Leuka. We also thank member laboratories of the United Kingdom Cancer Cytogenetic Group (UKCCG) for providing cytogenetic data and material. We are grateful to all the members of the NCRI Haematological Oncology Adult ALL Subgroup and the NCRI Childhood Cancer and Leukaemia Group (CCLG) Leukaemia Subgroup. Primary childhood leukaemia samples used in this study were provided by the Bloodwise Childhood Leukaemia Cell Bank working with the laboratory teams in the Bristol Genetics Laboratory, Southmead Hospital, Bristol: Molecular Biology Laboratory, Royal Hospital for Sick Children, Glasgow: Molecular Haematology Laboratory, Royal London Hospital, London: Molecular Genetics Service and Sheffield Children's Hospital, Sheffield. We also thank the Central England Haemato-Oncology Research Biobank for providing

patient samples. Finally, we thank all the clinicians who entered patients into the trial and the patients and families who agreed to take part.

Competing interests

All authors have no conflicts of interest to disclose

Data availability

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

1. Moorman AV. The clinical relevance of chromosomal and genomic abnormalities in B-cell precursor acute lymphoblastic leukaemia. Blood Reviews. 2012;26(3):123-35.
2. Wiemels JL, Cazzaniga G, Daniotti M, Eden OB, Addison GM, Masera G, et al. Prenatal origin of acute lymphoblastic leukaemia in children. The Lancet. 1999;354(9189):1499-503.
3. Greaves MF, Maia AT, Wiemels JL, Ford AM. Leukemia in twins: lessons in natural history. Blood. 2003;102(7):2321-33.
4. Maia AT, van der Velden VHJ, Harrison CJ, Szczepanski T, Williams MD, Griffiths MJ, et al. Prenatal origin of hyperdiploid acute lymphoblastic leukemia in identical twins. Leukemia. 2003;17(11):2202-6.
5. Mori H, Colman SM, Xiao Z, Ford AM, Healy LE, Donaldson C, et al. Chromosome translocations and covert leukemic clones are generated during normal fetal development. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(12):8242-7.
6. Wiemels JL, Ford AM, Van Wering ER, Postma A, Greaves M. Protracted and Variable Latency of Acute Lymphoblastic Leukemia AfterTEL-AML1 Gene Fusion In Utero. Blood. 1999;94(3):1057-62.
7. Den Boer ML, van Slegtenhorst M, De Menezes RX, Cheok MH, Buijs-Gladdines JG, Peters ST, et al. A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. The Lancet Oncology. 2009;10(2):125-34.
8. Mullighan CG, Su X, Zhang J, Radtke I, Phillips LAA, Miller CB, et al. Deletion of IKZF1 and Prognosis in Acute Lymphoblastic Leukemia. New England Journal of Medicine. 2009;360(5):470-80.
9. Roberts KG, Li Y, Payne-Turner D, Harvey RC, Yang YL, Pei D, et al. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. New England Journal of Medicine. 2014;371(11):1005-15.
10. Russell LJ, Capasso M, Vater I, Akasaka T, Bernard OA, Calasanz MJ, et al. Deregulated expression of cytokine receptor gene, CRLF2, is involved in lymphoid

transformation in B-cell precursor acute lymphoblastic leukemia. Blood. 2009;114(13):2688-98.
11. Mullighan CG, Collins-Underwood JR, Phillips LA, Loudin MG, Liu W, Zhang J, et al. Rearrangement of CRLF2 in B-progenitor- and Down syndrome-associated acute lymphoblastic leukemia. Nat Genet. 2009;41(11):1243-6.
12. Chapiro E, Russell L, Lainey E, Kaltenbach S, Ragu C, Della-Valle V, et al. Activating mutation in the TSLPR gene in B-cell precursor lymphoblastic leukemia. Leukemia. 2010;24(3):642-5.
13. Moorman AV, Schwab C, Ensor HM, Russell LJ, Morrison H, Jones L, et al. IGH@ Translocations, CRLF2 Deregulation, and Microdeletions in Adolescents and Adults With Acute Lymphoblastic Leukemia. Journal of Clinical Oncology. 2012;30(25):3100-8.
14. Schwab CJ, Jones LR, Morrison H, Ryan SL, Yigittop H, Schouten JP, et al. Evaluation of multiplex ligation-dependent probe amplification as a method for the detection of copy number abnormalities in B-cell precursor acute lymphoblastic leukemia. Genes Chromosomes Cancer. 2010;49(12):1104-13.
15. Buitenkamp TD, Pieters R, Gallimore NE, van der Veer A, Meijerink JPP, Beverloo HB, et al. Outcome in children with Down's syndrome and acute lymphoblastic leukemia: role of IKZF1 deletions and CRLF2 aberrations. Leukemia. 2012;26(10):2204-11.
16. Roberts KG, Morin RD, Zhang J, Hirst M, Zhao Y, Su X, et al. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. Cancer Cell. 2012;22(2):153-66.
17. Morak M, Attarbaschi A, Fischer S, Nassimbeni C, Grausenburger R, Bastelberger S, et al. Small sizes and indolent evolutionary dynamics challenge the potential role of P2RY8-CRLF2–harboring clones as main relapse-driving force in childhood ALL. Blood. 2012;120(26):5134-42.
18. Schwartzman O, Savino AM, Gombert M, Palmi C, Cario G, Schrappe M, et al. Suppressors and activators of JAK-STAT signaling at diagnosis and relapse of acute lymphoblastic leukemia in Down syndrome. Proceedings of the National Academy of Sciences. 2017;114(20):E4030-E9.
19. Wolman SR. Cytogenetic heterogeneity: its role in tumor evolution. Cancer Genet Cytogenet. 1986;19(1-2):129-40.
20. Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature. 2011;469(7330):356-61.
21. Ma X, Edmonson M, Yergeau D, Muzny DM, Hampton OA, Rusch M, et al. Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. Nature communications. 2015;6:6604.
22. Papaemmanuil E, Rapado I, Li Y, Potter NE, Wedge DC, Tubio J, et al. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. Nat Genet. 2014;46(2):116-25.
23. Potter NE, Ermini L, Papaemmanuil E, Cazzaniga G, Vijayaraghavan G, Titley I, et al. Single-cell mutational profiling and clonal phylogeny in cancer. Genome Research. 2013;23(12):2115-25.
24. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. Nature. 2011;472(7341):90-4.
25. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. New England Journal of Medicine. 2012;366(10):883-92.
26. Russell LJ, Jones L, Enshaei A, Tonin S, Ryan SL, Eswaran J, et al. Characterisation of the genomic landscape of CRLF2-rearranged acute lymphoblastic leukemia. 2017;56(5):363-72.
27. le Viseur C, Hotfilder M, Bomken S, Wilson K, Rottgers S, Schrauder A, et al. In childhood acute lymphoblastic leukemia, blasts at different stages of immunophenotypic maturation have stem cell properties. Cancer Cell. 2008;14(1):47-58.

28. Jeffries SJ, Jones L, Harrison CJ, Russell LJ. IGH@ translocations co-exist with other primary rearrangements in B-cell precursor acute lymphoblastic leukemia. Haematologica. 2014;99(8):1334-42.
29. Furness CL, Mansur MB, Weston VJ, Ermini L, van Delft FW, Jenkinson S, et al. The subclonal complexity of STIL-TAL1+ T-cell acute lymphoblastic leukaemia. Leukemia. 2018.
30. De Bie J, Demeyer S, Alberti-Servera L, Geerdens E, Segers H, Broux M, et al. Single-cell sequencing reveals the origin and the order of mutation acquisition in T-cell acute lymphoblastic leukemia. Leukemia. 2018.
31. Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V, et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature. 2014;506(7488):328-33.
32. Dussiau C, Fontenay M. Mechanisms underlying the heterogeneity of myelodysplastic syndromes. Exp Hematol. 2018;58:17-26.
33. Kim T, Tyndel MS, Kim HJ, Ahn JS, Choi SH, Park HJ, et al. The clonal origins of leukemic progression of myelodysplasia. Leukemia. 2017;31(9):1928-35.
34. Kent DG, Ortmann CA, Green AR. Effect of mutation order on myeloproliferative neoplasms. The New England journal of medicine. 2015;372(19):1865-6.
35. Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA, et al. Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. Science. 2008;322(5906):1377-80.
36. van Delft FW, Horsley S, Colman S, Anderson K, Bateman C, Kempski H, et al. Clonal origins of relapse in ETV6-RUNX1 acute lymphoblastic leukemia. Blood. 2011;117(23):6247-54.
37. Ford AM, Mansur MB, Furness CL, van Delft FW, Okamura J, Suzuki T, et al. Protracted dormancy of pre-leukemic stem cells. Leukemia. 2015;29(11):2202-7.
38. Melchor L, Brioli A, Wardell CP, Murison A, Potter NE, Kaiser MF, et al. Single-cell genetic analysis reveals the composition of initiating clones and phylogenetic patterns of branching and parallel evolution in myeloma. Leukemia. 2014;28(8):1705-15.
39. Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat Genet. 2014;46(3):225-33.
40. Davies NJ, Kwok M, Gould C, Oldreive CE, Mao J, Parry H, et al. Dynamic changes in clonal cytogenetic architecture during progression of chronic lymphocytic leukemia in patients and patient-derived murine xenografts. Oncotarget. 2017;8(27):44749-60.
41. Belderbos ME, Koster T, Ausema B, Jacobs S, Sowdagar S, Zwart E, et al. Clonal selection and asymmetric distribution of human leukemia in murine xenografts revealed by cellular barcoding. Blood. 2017;129(24):3210-20.
42. Schmitz M, Breithaupt P, Scheidegger N, Cario G, Bonapace L, Meissner B, et al. Xenografts of highly resistant leukemia recapitulate the clonal composition of the leukemogenic compartment. Blood. 2011;118(7):1854-64.
43. Shlien A, Campbell BB, de Borja R, Alexandrov LB, Merico D, Wedge D, et al. Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. Nat Genet. 2015;47(3):257-62.
44. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell. 2011;144(1):27-40.
45. Hart JR, Zhang Y, Liao L, Ueno L, Du L, Jonkers M, et al. The butterfly effect in cancer: a single base mutation can remodel the cell. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(4):1131-6.
46. Lee RS, Stewart C, Carter SL, Ambrogio L, Cibulskis K, Sougnez C, et al. A remarkably simple genome underlies highly malignant pediatric rhabdoid cancers. J Clin Invest. 2012;122(8):2983-8.
47. Cabelof DC, Patel HV, Chen Q, van Remmen H, Matherly LH, Ge Y, et al. Mutational spectrum at GATA1 provides insights into mutagenesis and leukemogenesis in Down syndrome. Blood. 2009;114(13):2753-63.

48. Hertzberg L, Vendramini E, Ganmore I, Cazzaniga G, Schmitz M, Chalker J, et al. Down syndrome acute lymphoblastic leukemia, a highly heterogeneous disease in which aberrant expression of CRLF2 is associated with mutated JAK2: a report from the International BFM Study Group. Blood.115(5):1006-17.

538 **Figure Legends**

539 **Figure 1. Multiple colour FISH showing examples of *IGH-CRLF2* as a clonal event**
540 **arising (A) first or (B) together with other tracked abnormalities (C) *P2RY8-CRLF2* as**
541 **a clonal event arising first or (D) together with other tracked abnormalities.**

542 • Representative FISH images on the right show examples of each leukaemic
543 sub-clone.

544 • Leukaemic sub-clone percentages for the diagnostic samples are indicated
545 next to each clone and only include populations above the 8% cut off.

546 • Dotted clone with "?": presumed but undetected founder clone.

547 **(A) For patient 23394** a total of 245 abnormal nuclei at diagnosis were scored for
548 probes to *CRLF2* (red/green) and *CDKN2A/B* (gold). In 34% of nuclei the *IGH-CRLF2*
549 translocations was an early event observed in the presence of two copies of
550 *CDKN2A/B*. This clone evolves into two sub-clones: one gains an extra copy of *IGH* in
551 17% of nuclei; one loses a single copy of *CDKN2A/B* in 49% of nuclei.

552 **(B) For patient 4001** a total of 129 abnormal nuclei at diagnosis were scored for probes
553 to *CRLF2* (red/green), *IKZF1* (gold) and *PAX5* (aqua). In 67% of nuclei the *IGH-CRLF2*
554 translocation was present with one copy of *IKZF1* and two copies of *PAX5*. This clone
555 evolves with loss of a single copy of *PAX5* in 33% of nuclei.

556 **(C) For patient 10924** a total of 227 abnormal nuclei at diagnosis were scored for
557 probes to *CRLF2* (red/green) and *IKZF1* (gold). In 11% of nuclei the *P2RY8-CRLF2* fusion
558 is an early event before the loss of a single copy of *IKZF1* observed in the bulk
559 leukaemic clone at 89%.

560 **(D) For patient 322** a total of 70 abnormal nuclei at diagnosis were scored for probes
561 to *CRLF2* (red/green) and *CDKN2A/B* (gold). In 67% of nuclei the *P2RY8-CRLF2* fusion

and loss of a single copy of *CDKN2A/B* present together. A further 33% of cells gained an additional copy of the *CRLF2* probe, which usually indicates the presence of an additional sex chromosome (+X/Y).

Figure 2. Multiple colour FISH showing examples of the sub-clonal nature of (A) *P2RY8-CRLF2* (B) *IGH-CRLF2*

- Representative FISH images on the right show examples of each leukaemic sub-clone.
- DAPI has been removed from panel B in order for the *PAX5* aqua signals to be observed.
- Leukaemic sub-clone percentages for the diagnostic samples are indicated next to each clone and only include populations above the 8% cut off.
- Dotted clone with "?": presumed but undetected founder clone.

(A) For patient 3789 a total of 144 abnormal nuclei at diagnosis were scored for probes to *CRLF2* (red/green) and *IKZF1* (gold). In 11% of nuclei both copies of *IKZF1* are deleted. This clone subsequently acquires the *IGH-CRLF2* translocation in a further 83% of cells.

(B) For patient 2017 a total of 211 abnormal nuclei at diagnosis were scored for probes to *CRLF2* (red/green) and *CDKN2A/B* (gold) and *PAX5* (aqua). The earliest abnormal clone detected has loss of a single copy of *CDK2NA/B* and *PAX5* suggesting *P2RY8-CRLF2* is a sub-clonal event. This clone acquired *P2RY8-CRLF2* in a further 70% of nuclei.

Figure 3. Multiple colour FISH showing examples of parallel evolution in (A-E) *P2RY8-CRLF2* (F) *IGH-CRLF2* patients driven by reiterative CNAs within the same genes

- Leukaemic sub-clone percentages for the diagnostic samples are indicated next to each clone and only include sub-clones above the 8% cut off.
- Dashed lines present possible alternate routes to final sub-clone. It is likely that one sub-clone was not competitive enough and the other acquired the additional abnormality to continue evolving.
- Dotted clone with "?": presumed but undetected founder clone.

(A) For patient 12200 a total of 269 abnormal nuclei at diagnosis were scored for probes to *CRLF2* and *CDKN2A/B*. Two independent clones evolved from an undetectable founder clone either losing both copies of *CDKN2A/B* in 19% or acquiring the *P2RY8-CRLF2* fusion in 16%. The major clone encompassing 65% of nuclei was observed to have both abnormalities from the previous two independent clones with an additional copy of *CRLF2*, which usually indicates the presence of an additional sex chromosome.

(B) For patient 3173 a total of 58 abnormal nuclei at diagnosis were scored for probes to *CRLF2* and *IKZF1.* Two independent clones evolved from an undetectable founder clone either losing a single copy of *IKZF1* in 10% or the formation of *P2RY8-CRLF2* in 45%. A clone was observed to have evolved from either previous clone by gaining the *P2RY8-CRLF2* fusion or losing a single copy of *IKZF1*, respectively.

(C) For patient 4954 a total of 50 abnormal nuclei at diagnosis were scored for probes to *CRLF2*, *IKZF1* and *CDKN2A/B*. The earliest clone detected harboured the *P2RY8-CRLF2* fusion with loss of a single copy of *CDKN2A/B* in 18%. Two independent clones evolved from this population either losing a single copy of *IKZF1* in 30% or losing the

second copy of *CDKN2A/B* in 20%. The major clone encompassing 32% of nuclei was observed to have evolved from either previous clone by losing a second copy of *CDKN2A/B* or losing a single copy of *IKZF1*, respectively.

(D) For patient 20753 a total of 50 abnormal nuclei at diagnosis were scored for probes to *CRLF2*, *IKZF1* and *PAX5*. The earliest abnormal clone detected harboured the *P2RY8-CRLF2* fusion with loss of a single copy of *IKZF1* in 18% of nuclei. Two independent clones evolved from this population either losing the second copy of *IKZF1* in 18% or doubling up the derived chromosome X to give two copies of *P2RY8-CRLF2* in 20%. This latter clone evolves further to lose a single copy of *PAX5* in 20% and the second copy of *IKZF1* in a further 24%.

(E) For patient 5817 a total of 120 abnormal nuclei at diagnosis were scored for probes to *CRLF2* and *IKZF1*. Two independent clones evolved from an undetectable founder clone. A signal pattern indicative of a rearrangement of *CRLF2* (1R1G1F) in the presence of a single copy of *IKZF1* was observed in 15% of cells. *P2RY8-CRLF2* alone (FISH signal pattern 1R0G1F) was observed in an independent clone of 28%. This clone evolved to lose a single copy of *IKZF1* in 52% followed by doubling up the derived sex chromosome to give two copies of *P2RY8-CRLF2* in 21%.

(F) For patient 21572 a total of 152 abnormal nuclei at diagnosis were scored for probes to *CRLF2*, *IKZF1* and *PAX5*. Two independent clones evolved from an undetectable founder clone with different signal patterns for the *CRLF2* rearrangements. A signal pattern indicative of a rearrangement of *CRLF2* (1R1G1F) was observed in 13% of cells in the presence of a single copy of *IKZF1*. *IGH-CRLF2* alone was observed in an independent clone of 13%. The signal pattern observed in these cells was 1R0G1F, which in conjunction with a split *IGH* FISH signal pattern, suggests

the cells have an *IGH-CRLF2* translocation and a deletion on the derived sex chromosome involved in the translocation. This clone evolves to lose a single copy of *IKZF1* in 43% followed by loss of a single copy of *PAX5* in a further 31%.

Figure 4. Sub-clonal architecture demonstrated by single cell analysis

- Sub-clone percentages for the diagnostic samples are indicated next to each clone.
- The blue triangle and orange square denote groups of clonal abnormalities including SNVs and CNAs.
- Those sub-clones identified in the primary mouse spleen (S) and bone marrow (BM) are indicated using black arrows. Where secondary mice were tested; S-S indicates spleen cells and S-BM, bone marrow cells.
- Where stated, alterations were tracked by either patient specific fusion and mutation assays or generic copy number assays.
- Gene copy numbers, where appropriate are indicated next to the gene name.

(A) Non-DS-ALL patient 11543 - In the diagnostic sample 310 cells were successfully screened. Bulk diagnostic cells engrafted in three mice, the spleen and bone marrow (250 cells from each) were investigated from one mouse. At diagnosis the *IGH-CRLF2* translocation is observed on a background of clonal abnormalities (including loss of *IKZF1*) in 40% of cells with disruption of the second copy of *IKZF1* in a further 48% of cells. No abnormalities were detected in 8% of cells. ^ Indicates both copies of *IKZF1* are disrupted. All clones were observed in cells isolated from the spleen with the two most frequent clones being observed in the bone marrow.

(B) Non-DS-ALL patient 21819 - In the diagnostic sample 296 cells were successfully screened. Bulk diagnostic cells engrafted in three mice and were serially transplanted. We investigated the spleen and bone marrow from one primary mouse (229 and 237 cells respectfully) and one secondary mouse (76 and 221 cells respectfully). This leukaemia presented at diagnosis with a linear sub-clonal architecture characterised

by multiple aberrations; the deletion of *IKZF1* was sub-clonal (96% of cells). The bulk sub-clone represented 89% of cells and harboured a *P2RY8-CRLF2* fusion. This sub-clone was observed across primary and secondary mice (S) in cells harvested from both the spleen and bone marrow. Primary mice also harboured the smaller sub-clone.

(C) DS-ALL patient 19599 - In the diagnostic sample 257 cells were successfully screened. Bulk diagnostic cells engrafted in three mice, the spleen and bone marrow (229 and 237 cells respectively) were investigated from one mouse. At diagnosis three independent subpopulations evolved from a backbone of clonal aberrations (which included loss of *IKZF1* and a *IGH-CRLF2* translocation). A small clone observed in 2% of cells acquired biallelic loss of *IKZF1*. Two further clones acquired unique independent *IKZF1* mutations each then fostering additional alterations; one losing a single copy of *BTLA* and gaining a *TOP3A* mutation (13% of cells) followed by the acquisition of a *JAK2* mutation (54% cells); the other acquired another independent *JAK2* mutation (4% of cells) followed by a *GNB1* mutation (27%). This was the only clone to be detected in the mouse spleen and bone marrow cells from the mouse. ^ In those cells with an *IKZF1* mutation, only the mutant signal was present, confirming *IKZF1* copy number loss and mutation of the remaining copy.

(D) DS-ALL patient 11538 - In the diagnostic sample 324 cells were successfully screened. Bulk diagnostic cells engrafted in three mice, the spleen and bone marrow (243 and 240 cells respectively) were investigated from one mouse. At diagnosis two independent clones evolved from a very complex backbone of aberrations (including loss of *IKZF1* and a *P2RY8-CRLF2* fusion). A small clone was characterised by disruption of the second copy of *IKZF1* alone (7% of cells) compared to the bulk clone which

729 acquired 5 further mutations and then disruption of the second copy of *IKZF1* (57% of
730 cells). ^ Indicates both copies of *IKZF1* are disrupted. All clones were observed in the
731 cells harvested from mouse spleen and bone marrow. The small clone characterised
732 by disruption of the second *IKZF1* allele was the major clone appearing in the mice.
733

Figure 1. Multiple colour FISH showing examples of *IGH-CRLF2* as a clonal event arising (A) first or (B) together with other tracked abnormalities (C) *P2RY8-CRLF2* as a clonal event arising first or (D) together with other tracked abnormalities.

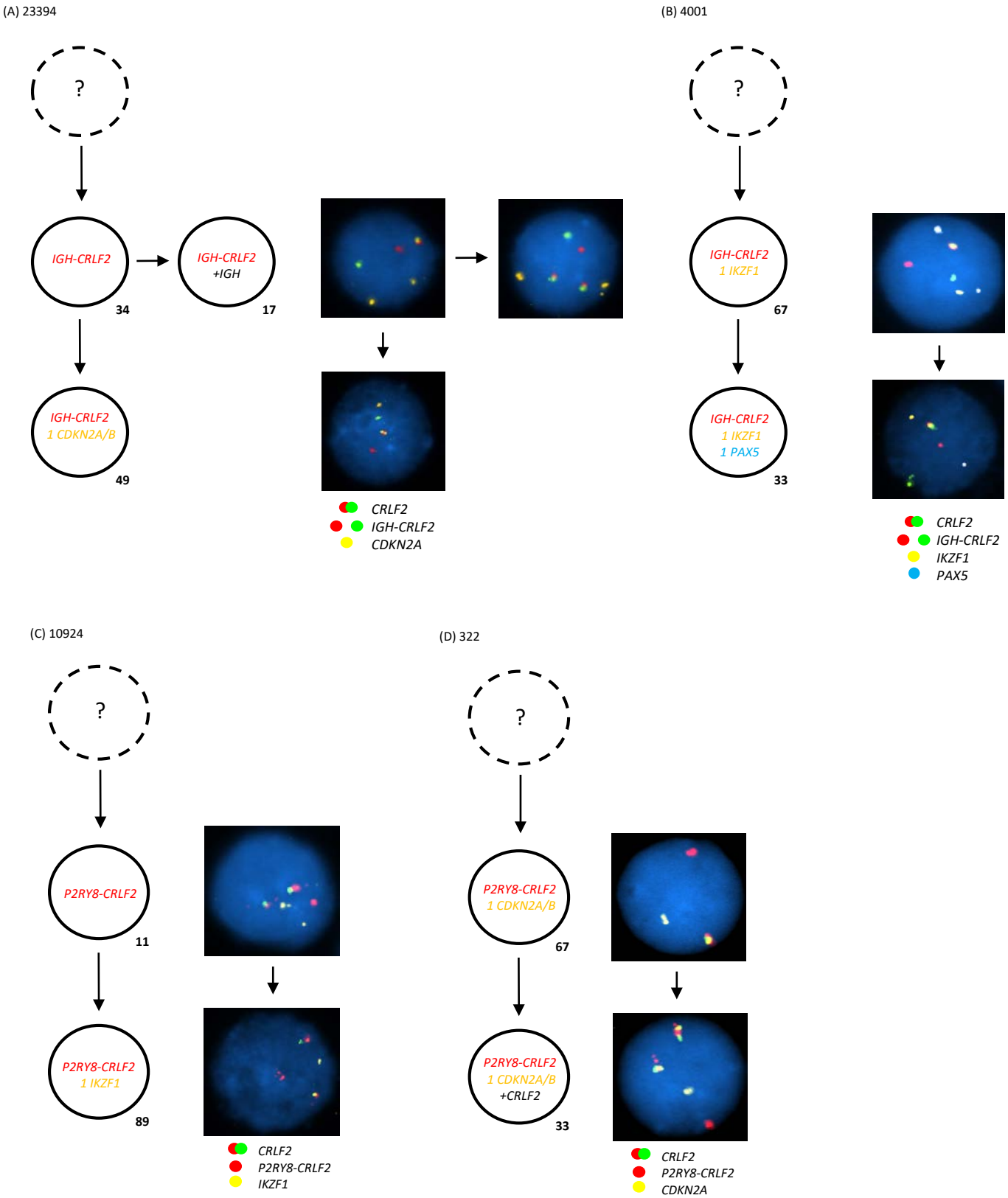

Figure 2. Multiple colour FISH showing examples of the sub-clonal nature of (A) *P2RY8-CRLF2* (B) *IGH-CRLF2*

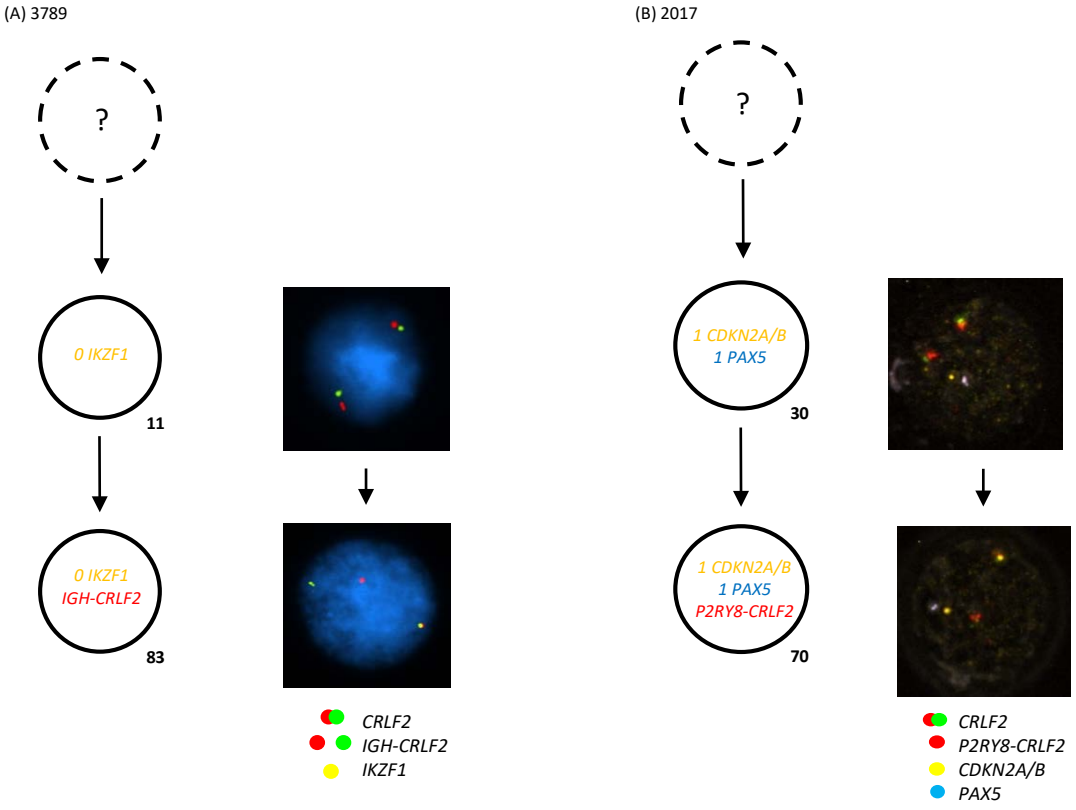

Figure 3. Multiple colour FISH showing examples of parallel evolution in (A-E) *P2RY8-CRLF2* (F) *IGH-CRLF2* patients driven by reiterative CNAs within the same genes

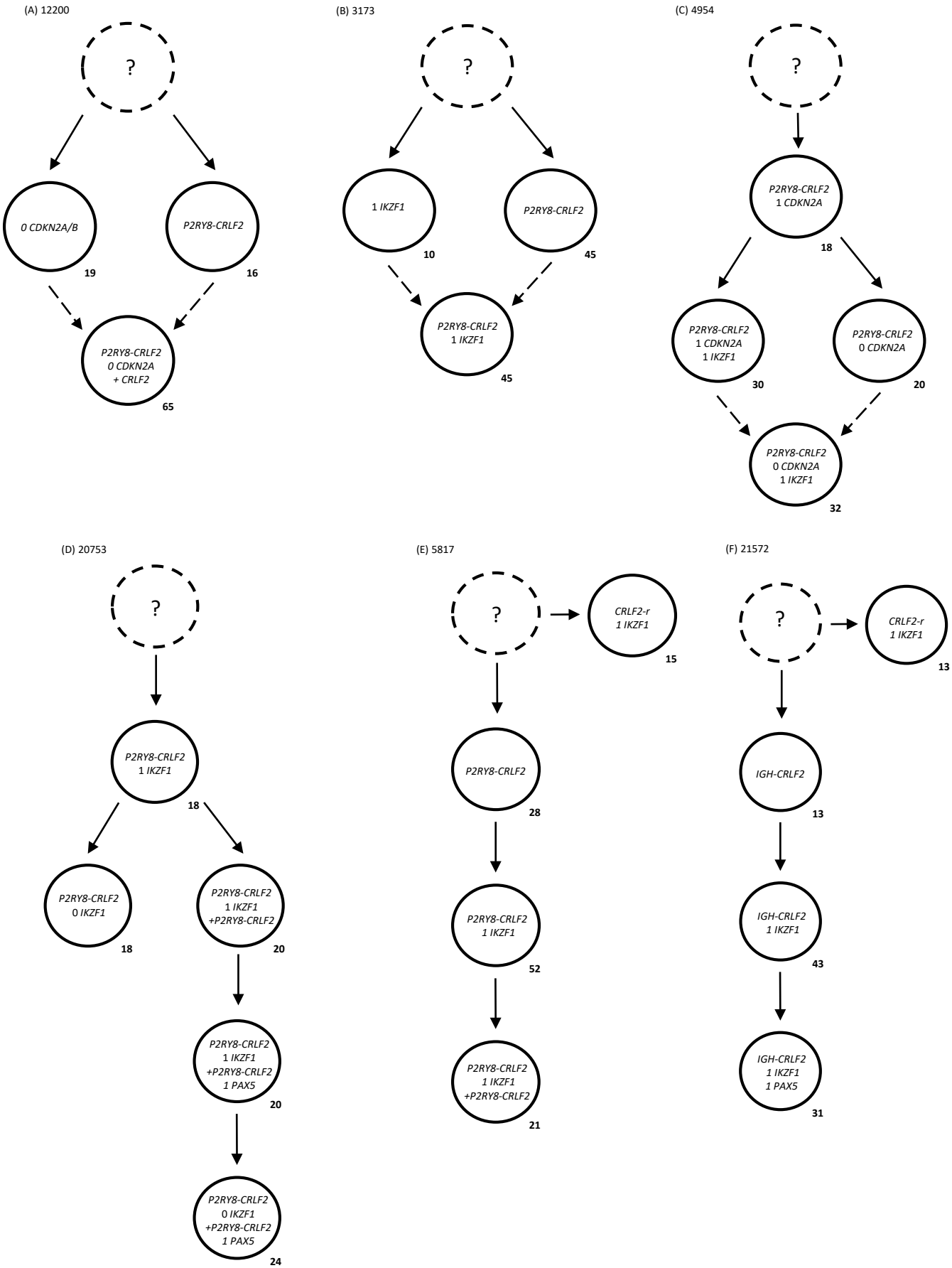

Figure 4. Sub-clonal architecture demonstrated by single cell analysis