Complex genotype-phenotype relationships shape the response to treatment of Down Syndrome Childhood Acute Lymphoblastic Leukaemia

Christoph Lutz[1,7,8*], Virginia A. Turati[1*], Ruth Clifford[7], Petter S. Woll[2,3,10], Thomas Stiehl[4], Anders Castor[5], Sally A. Clark[2,3], Helen Ferry[2,3], Veronica Buckle[2] Andreas Trumpp[6], Anthony Ho[7], Anna Marciniak-Czochra[4], Javier Herrero[1], Anna Schuh[9], Sten-Eirik W. Jacobsen[2,3,10] and Tariq Enver[1,11]

[1]University College London, Cancer Institute, London WC1E 6BT, UK
[2]MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine,
[3]University of Oxford, Oxford, OX3 9DS, UK
[3]Haematopoietic Stem Cell Biology, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, OX3 9DS, UK
[4]Institute of Applied Mathematics Mathematikon, Heidelberg University, Germany
[5] Department of Paediatric Haematology/Oncology, Skane University Hospital, Sweden
[6] Deutsches Krebsforschungszentrum Heidelebrg (DKFZ), Germany
[7] Department of Internal Medicine V, Heidelberg University, Germany
[8]Praxis für Onkologie, Neversstrasse 5, 56077 Koblenz, Germany
[9]Department of Haematology, Oxford Radcliffe Hospitals NHS Trust, Oxford OX3 9DU, UK
[10]Center for Hematology and Regenerative Medicine, Karolinska Institutet and Karolinska University Hospital Huddinge, Stockholm, Sweden
[11] Division of Molecular Medicine and Gene Therapy, Lund University, A12, 221 84 Lund Lund, Sweden
[*] These authors contributed equally to this work

Correspondence to:

Tariq Enver PhD
UCL Cancer Institute
Stem Cell Laboratory
Paul O'Gorman Building
72 Huntley Street
London WC1E 6BT

email: t.enver@cancer.ucl.co.uk
phone: +44 (0)20 7679
Fax: +44 (0)20 7679 6914

Abstract

Extensive genetic and epigenetic variegation has been demonstrated in many malignancies. Importantly, their interplay has the potential to contribute to disease progression and treatment resistance. To shed light on the complex relationships between these different sources of intra-tumour heterogeneity, we explored their relative contributions to the evolutionary dynamics of Acute Lymphoblastic Leukemia (ALL) in children with Down syndrome, which has particularly poor prognosis. We quantified the tumour propagating potential of genetically distinct sub-clones using serial transplantation assays and SNP-arrays. While most leukaemias were characterized by a single dominant subclone, others were highly heterogeneous. Importantly, we provide clear and direct evidence that genotypes and phenotypes with functional relevance to leukemic progression and treatment resistance can co-segregate within the disease. Hence, individual genetic lesions can be restricted to well-defined cell immunophenotypes, corresponding to different stages of the leukemic differentiation hierarchy and varied proliferation potentials. As a result of this difference in fitness, which can be accurately quantified via competitive transplantation assays, matching diagnostic, post-treatment, and relapse leukemias can be dominated by different genotypes, including pre-leukemic clones persisting throughout the disease progression and treatment. Intriguingly, plasticity also appears to be a temporally defined property that can segregate with genotype. These results suggest that Down Syndrome ALL should be viewed as a complex matrix of cells exhibiting genetic and epigenetic heterogeneity that foster extensive clonal evolution and competition. Therapeutic intervention reshapes this 'eco-system' and may provide the right conditions for the preferential expansion of selected compartments and subsequently relapse.

Introduction:

The existence of intra-tumour genetic heterogeneity, both genetic and epigenetic, has been documented in both solid and haematological malignancies[1–3]. In studies of childhood acute lymphoblastic leukaemia, we and others previously used multicolour-FISH and whole genome sequencing to show that these tumours are clonally variegated, and to demonstrate through serial transplantation assays, that leukaemia tumour propagating cells (TPC) are also genetically heterogeneous[4,5]. This finding was supported by our later genetic analysis of immunophenotypically distinct ALL diagnostic subpopulations which demonstrated similar distribution of subclones within a disease-specific primitive B-cell compartment (StemB, $34^{+}38^{-}19^{+}$) and the disease bulk in most, but, crucially, not all leukaemias[6]. Such

experiments relied on high-depth bulk whole-genome sequencing followed by single-cell qPCR and, together with our earlier work, suggested that genotype and phenotypes related to differentiation stage and cell-cycle state might not necessarily co-segregate within the disease. Importantly, as chemotherapy-induced selection acts on these same phenotypic traits[6–8], this lack of segregation also underpinned the survival of most genotypes to the early stages of childhood ALL treatment (namely, induction chemotherapy – where maximum cytotoxicity is achieved); as we demonstrated using a xenograft model of transplantation and treatment of primary leukaemias[6].

Large-scale comparisons of matching diagnostic and relapse samples from paediatric ALL patients, have nonetheless previously reported that in some patients a subset of diagnostic subclones, or of their later evolutionary descendants, can be selectively enriched at the time of disease recurrence[9,10]. Assuming that these clones did not become dominant as a result of stochastic selection, this finding opens to the possibility that fitness-based genotypic selection might potentially occur on a longer timescale than that of induction chemotherapy. It is not unreasonable to speculate that in a treatment protocol that lasts up to four years even subtle differences in fitness might eventually lead to clonal dominance. Another possibility is that in certain settings, genotype, and treatment-associated phenotypes do, in fact, segregate (either already at diagnosis or later because of further genomic evolution). In support of this latter hypothesis, our analysis of $34^{+}38^{-}19^{+}$ (StemB) ALL cells indicated that in a small proportion of leukaemias (1 out of 5 analysed), specific genotypes might indeed be preferentially enriched within developmentally more primitive and quiescent disease compartments[6]. Furthermore, recent evidence suggests thar mutations induced by chemotherapy treatment itself (particularly those targeting p53) might be responsible for - post-diagnosis – generating new, fitter subclones, and inducing drug-resistance; therefore, leading to relapse[11].

To shed light on the complex genotype-phenotype relationship that shape ALL, we here conduct a comprehensive investigation into the relative contribution of different sources of intratumor heterogeneity, namely genetic makeup and epigenetically determined properties - particularly cell cycle state, TPC activity and differentiation stage - to the disease initiation, sustenance, and treatment resistance. In doing so we focused our attention Down's syndrome associated ALL as this rarer disease subtype has so far remained poorly characterised while being associated with considerably worse prognosis than ALL in children without Down syndrome[12,13].

Altogether our results indicate that ALL should be viewed as a complex matrix of cells exhibiting genetic and epigenetic heterogeneity. Our work provides evidence for, in some patients, diversification in functional properties amongst immunophenotypically-distinct leukemic subpopulations. In these setting, those rare leukemic cells which remain after induction chemotherapy, and primarily account for minimal residual disease (MRD), are highly enriched for the stem/B phenotype; suggesting that the more primitive and less proliferative leukemic compartments likely constitute the pool of cells from which relapse arises. Crucially, in this setting where genotype and resistance-associated phenotypes co-segregate, bottleneck selection at phenotype level is accompanied by evidence of a clonal sweep.

Methods:

Patient samples

Diagnostic and follow-up ALL bone marrow (BM) samples were obtained after informed consent and the approval of the relevant research ethics committee from patients at the Paediatrics Haematology Unit, Lund University Hospital, Sweden. Samples from 4 children with Down's ALL were analyzed. For further patient characteristics see Table 1.

Cell separation, phenotyping, and sorting

Total mononuclear cells (MNCs) were isolated by ficoll gradient centrifugation and directly cryopreserved in DMSO for later use. In some cases, $CD34^{+}$ cells were enriched by magnetic bead separation (StemCell Technologies or Miltenyi). After thawing, dead cells were evaluated and excluded by FACS after staining with DAPI. CD34-enriched cells or MNCs from BM were stained with anti-CD19 PE (BD-Pharmingen), CD34 FITC (BD-Pharmingen) and CD38 APC (BD-Pharmingen). Cells were analyzed in staining buffer containing DAPI at 0·1µg/ml, for live cell analysis. I.) HSC ($34^{+}38^{-}19^{-}$); II.) Stem/B ($34^{+}38^{-}19^{+}$) III.) Progenitor ($34^{+}38^{+}19^{-}$) IV.) ProB ($34^{+}38^{+}19^{+}$) and V.) PreB/Mature B-cells ($34^{-}19^{+}$) were purified by flow cytometry (FACS Aria, BD-Pharmingen). Data acquisition and analysis were done with CellQuest (Beckon Dickinson) or FlowJo (Tree star) software.

Cell cycle analysis of leukaemia subpopulations

Cells were stained with CD19-PECy5 (BioLegend), then fixed and permeabilized with 1.6% paraformaldehyde and 90% ice-cold methanol. This was followed by staining with CD34-APC (BD), CD38-PETxR (Invitrogen) and Ki67-FITC (Becton-Dickinson). For DNA content analysis cells were incubated with 0·5µg/mL DAPI. Stained cells were analyzed by excitation

143 of DAPI with a violet laser on a FACS LSRII SORP (Becton-Dickinson) at the University of
144 Oxford (UK).
145
146 **FISH analysis**
147 Cells were fixed on slides in methanol/acetic acid fixative (3:1 vol/vol) and then hybridized
148 with BCR-ABL1 FISH probes using the LSI BCR/ABL1 Dual Color Dual Fusion Translocation
149 probes (Abbot). The slides were analyzed with an Olympus BX51 microscope equipped with
150 epifluorescence and a triple band pass filter. Images were captured by using a Sensys
151 charge-coupled device camera (Photometrics, Tucson, AZ) and MacProbe software (Applied
152 Imaging, Newcastle upon Tyne, U.K.). When possible, and unless otherwise specified at
153 least 100-150 nuclei were analyzed in each sample
154
155 **SNP-Array**
156 SNP-Arrays we performed as previously describe in Schuh et al., 2012[14]
157
158 **NSG-mouse transplantation assay**
159 Primary childhood ALL cells were transplanted into 8-12 weeks old NOD/SCID IL2Rγ^{hull}
160 (NSG) sub-lethally irradiated mice (either females or males) via intramedullary injection. To
161 minimize possible adverse effects of sublethal irradiation, mice were administered acid water
162 for a week prior to the procedure, and Baytril (resuspended at 25.5 mg/kg in the drinking
163 water) for the 2 weeks following it. Sub-lethal irradiation was achieved with a single dose of
164 375 cGy. Unless otherwise stated, each mouse received 2 x 10^5 primary leukemia cells
165 resuspended in 40µl PBS 0.5% FBS. In the case of secondary limiting dilution assays, a
166 specified equal dose of treated and control leukemic cells harvested from the BM of primary
167 recipients was injected. Twelve-week post-injection mice were sampled by bone marrow
168 aspiration and the percentage of human engraftment was evaluated by flow cytometry
169 (hCD45/(hCD45$^+$ + mCD45$^+$)). At the same time, human cells were also FACS sorted for
170 downstream applications. Mice displaying at least 70% human engraftment were then
171 randomly assigned to either control or treatment groups (see below for details on the
172 treatment protocol). At the end of the treatment window all mice were culled, and their tibias,
173 femurs, pelvises, spleen, and brain harvested, and cells were then stained for FACS sorting.
174
175 **Modeling of clonal evolution**
176 See supplementary materials.
177

178 **Small cell numbers RNA sequencing**

179 Equivalent cell numbers (400 cells per sample) were flow sorted directly into 800µl Trizol
180 reagent (Invitrogen) and snap frozen in dry ice (long term storage at -80C). At the time of
181 extraction, the samples were thawed at RT and 160ul of chloroform was added to each.
182 Following a centrifugation step the RNA was isolated from the aqueous phase and
183 precipitated through the addition of equal volumes of isopropanol supplemented with 20µg
184 linear polyacrylamide. Samples were washed twice in 80% ethanol (first wash over night at
185 4°C, second wash 5 minutes at RT). RNA pellets were resuspended in 3-15µl of
186 diethylpyrocarbonate treated water (DEPC). RNA was then quantified by loading of 0.5-1ul
187 on an Agilent Bionalyser RNA 6,000 pico chip. Where possible equivalent amounts of total
188 RNA (100pg) from all samples were used for first strand synthesis with the SmartERv3 kit
189 (Takara Clontech) followed by 15-18 cycles of amplification (according to manufacturers'
190 instruction). cDNA was purified on Agencourt AMPureXP magnetic beads, washed twice
191 with fresh 80% ethanol and eluted in 17µl elution buffer. 1µl cDNA was quantified with Qubit
192 dsDNA HS (Molecular Probes) and checked on an Agilent Bioanalyser high sensitivity DNA
193 chip. Sequencing libraries were produced from 150pg input cDNA using Illumina Nextera XT
194 library preparation kit. A 1:4 miniaturized version of the protocol was adopted (see "Fluidigm
195 Single-Cell cDNA Libraries for mRNA sequencing", PN_100-7168_L1). Tagmentation time
196 was 5mins, followed by 12 cycles of amplification using Illumina XT 24 or 96 index primer kit.
197 Libraries were then pooled (1-2ul per sample depending on the total number of samples)
198 and purified with equal volumes (1:1) of Agencourt AMPureXP magnetic beads. Final elution
199 was in 66-144ul of resuspension buffer (depending on the total number of pooled samples).
200 Libraries were checked on an Agilent Bioanalyser high sensitivity DNA chip (size range 150-
201 2000bp) and quantified by Qubit dsDNA HS (Molecular Probes). Libraries were sequenced
202 on Illumina® NextSeq 500 using 150bp paired end kits as per manufacturer's instructions.

203

204 **Bioinformatics**

205 Sequencing data was assessed to detect sequencing failures using FASTQC and lower
206 quality reads were filtered or trimmed using *TrimGalore*
207 (https://github.com/FelixKrueger/TrimGalore). Outlier samples containing low sequencing
208 coverage or high duplication rates were discarded.

209

210 **Processing of bulk RNAseq**

211 Bulk RNAseq samples were mapped to the human reference GRCh38 using *tophat2*
212 (https://ccb.jhu.edu/software/tophat/index.shtml). Analyses were performed within the R
213 statistical computing framework, version 3.5 using packages from *BioConductor version 3.7*
214 (https://Bioconductor.org). Data was combined into a per-gene count matrix using
215 featureCounts from the subread package. The *DEseq2* BioConductor package was used for
216 outlier detection, normalization and differential gene expression analyses. All downstream
217 analyses used Rlog transformed data.

218 **Data availability**

219 SNP-array and RNAseq data are available at EGA under accession number
220 EGAS00001005968.

221

222 **Results**

223 We here explored the role of intra-tumour heterogeneity in the progression and response to
224 treatment of Down's associated ALL. We initially focused on genetic variegation and
225 described the phylogenetic hierarchies of diagnostic material from 4 patients. Serial
226 transplantation of bulk ALL samples into immune-deficient NSG-mice was concomitantly
227 utilized to evaluate the tumour propagating cell activity (TPC) and dynamic behaviours of the
228 different subclones identified. SNP-arrays were performed on diagnostic specimens as well
229 as cells isolated from the BM of primary and secondary transplanted mice (see schemata in
230 Figure 1A). In all patients, but one, a single dominant clone was detected both at diagnosis
231 and in all primary and secondary transplanted mice (Figure 1B). In patient 2, instead, a more
232 complex picture was observed, with several subclones coexisting within the tumour and
233 contributing to a different extent to engraftment across recipients. In this patient we detected
234 two deletions, on chromosomes 9 and 14 respectively, as well as an amplification on
235 chromosome X (Figure 1C). Considering the mutational mosaicism, we observed in some of
236 the engrafted mice, these observations allowed re-construction of a simple clonal
237 phylogenetic architecture (Figure 1D, and Supp Material for Mosaicism data). However, a
238 higher-depth genetic analysis of chromosome 9 suggested a far more complicated clonal
239 landscape. A heterozygous chromosome 9 deletion was detected across engrafted animals
240 with different degrees of homozygous involvement: to the extreme of copy neutral loss of
241 heterozygosity (cnLOH) (Figure 2A). A simple explanation to this observation is that pre-
242 existing sub-clones present at diagnosis at sub-threshold levels, were expanded through
243 secondary transplantation until reaching detection levels. Alternatively, continuous
244 broadening of the chromosome 9 deletion could have occurred over the short time frame of
245 the serial transplantation assays. This latter hypothesis would be indicative of active ongoing

246 clonal evolution. Based on the mosaicism data, and assuming no new mutations developed
247 during transplantation, we were able to reconstruct a more comprehensive phylogenetic tree.
248 To do so, we developed a custom-made mathematical model which we used to calculate all
249 the theoretically possible diagnostic clone-combinations supported by our SNPs data
250 analysis (Supplementary Table 1). In total 15-29 diagnostic clones were necessary to
251 explain the mutational read-out from the xenografts. Of note, since SNP-arrays are limited in
252 both genomic resolution (i.e., deletion size) and sensitivity to low abundance copy number
253 variants (CNV) our model would if anything underestimate the number of clones (Figure 2B
254 and supplementary methods). Nonetheless, this finding clearly illustrates the high clonal
255 complexity of ALL and is line with our earlier single-cell WGS based characterization of
256 genetic heterogeneity in non-Down Syndrome ALL, which also identified branching clonal
257 architectures with up to 19 descendant subclones[6].
258
259 We subsequently sought to shed light on the phenotype-genotype relationship that defines
260 the properties of individual leukemic subclones, by performing SNP analysis on bulk tumour
261 cells and flow-sorted immunophenotypically-distinct ALL subpopulations from three Down's
262 syndrome ALL patients both before and after xenotransplantation (see schemata in Figure
263 3A). Analysis of samples from patients 2 and 3 (supplementary Figure 1) revealed identical
264 clonal architectures in diagnostic bulk, stem/B and proB-like cells for both patients (Figure
265 3BC). All samples analysed from patient 3 confirmed the existence of a single dominant
266 leukemic clone (Figure 3C and supplementary table 1). In addition to the founding
267 chromosome 21 amplification, a third patient (pt.4) we examined harboured a BCR-ABL
268 translocation. FISH analysis showed that in this leukemia the fusion gene marked proB- and
269 preB-like sub-compartments, but not cells with the more primitive and disease-specific
270 stem/B ($34^{+}38^{-}19^{+}$) phenotype[15] (Figure 4A). Notably, in serial transplantation experiments
271 flow-sorted stem/B cells only gave rise to grafts with stem/B like cells but not more
272 differentiated (proB-preB) B-cell phenotypes. These stem/B like grafts remained BCR-ABL-
273 negative. Similarly, proB cells derived grafts only contained proB and preB cells, all of which
274 retained the BCR-ABL-fusion. Collectively these data suggest lack of phenotypic plasticity
275 amongst distinct diagnostic cells from this patient.
276
277 We next compared the relative fitness of the two subpopulations (stem/B and proB) in a
278 competitive transplantation setting facilitated by (i) the presence of the BCR-ABL
279 translocation as a clonotypic marker and (ii) the non-interchangeability of their
280 immunophenotypes. We transplanted NSG-mice with different ratios of stem/B and proB-like
281 ALL cells and read out engraftment. In all cases the BCR-ABL^{+} ProB-like cells displayed

282 increased engraftment fitness, most clearly exemplified in the setting where proB cells
283 represented only 10% of the initial inoculum and yet constituted 95% of the resultant
284 leukaemia (Figure 4B). Reasoning that enhanced fitness may stem from a proliferative
285 advantage we next analysed the cell cycle status of engrafted cells. In all engrafted
286 recipients proB cells displayed a more active cycling status compared to their stem/B
287 counterparts, which had a significantly larger proportion of dormant (G_0) cells (Figure 4C and
288 Supplementary Figure 2).
289
290 We subsequently speculated that the different cycling properties of these compartments
291 could be expected to affect their relative sensitivity to therapy. To experimentally determine
292 how these different ALL sub-compartments behaved during therapy, we analysed leukemic
293 subpopulations from serially collected diagnosis, remission, and relapse specimens. ProB-
294 like ALL cells were indeed preferentially eradicated during chemotherapy. Indeed, while both
295 populations were shrank by treatment to similar frequency (0.01%), ProB cells initially
296 represented approximately 70% of leukaemic cells detected at diagnosis while stem/B-like
297 ALL cells were only 8% (Figure 5A). After induction chemotherapy extremely few proB-like
298 cells remained, none of which carried the BCR-ABL translocation. While the behaviour of
299 these subpopulations during treatment contrasts starkly with their fitness as assessed by
300 xenotransplantation, it is entirely consistent with the known relative vulnerability of cycling
301 versus dormant cells to chemotherapy. Similarly, to what has previously been observed in
302 the context of bacteria and antibiotics resistance, these data also support the idea of the
303 existence of a trade-off on phenotypic traits between environments with and without drug[16].
304 The reduced competitive ability of the less proliferative species (StemB cells) in the absence
305 of treatment hence becomes an advantage in the presence of chemotherapy, while the
306 highly proliferative ProB cells pay a cost to their initial fitness.
307
308 Interestingly, the relapse sample was populated by both stemB and proB-like cells (Figure
309 5A). However, in contrast to the situation at diagnosis leukemic proB cells present at relapse
310 were BCR-ABL-negative and both relapse stemB and proB cells showed phenotypic
311 plasticity upon serial transplantation (Figure 5B). This data is somewhat counterintuitive in
312 that it argues against the simplest origins of relapse from the predominant stemB or proB
313 clones present at diagnosis. The stem/B compartment at presentation lacked phenotypic
314 plasticity in that it was unable to regenerate proB- or preB-like cells. Since the proB cells at
315 relapse were BCR-ABL-negative, relapse could also not have been seeded by the
316 predominant diagnostic proB-like compartment that was positive for BCR-ABL.
317

318 Taking advantage of the xenograft system we further investigated the phenotypic properties
319 of diagnostic and relapse cells at the molecular level by performing small cell number bulk
320 RNA sequencing (RNAseq) on cells harvested from either primary and or secondary
321 xenografts. We hence characterised six distinct cell populations i) diagnostic stemB cells
322 from primary xenografts, ii) diagnostic proB cells from primary xenografts iii) relapse stemB
323 cells harvested from secondary xenografts injected with stemB cells isolated from primary
324 xenografts iv) relapse proB cells harvested from secondary xenografts injected with stemB
325 cells isolated from primary xenografts, v) relapse proB cells harvested from secondary
326 xenografts injected with proB cells isolated from primary xenografts, and vi) relapse stemB
327 cells harvested from secondary xenografts injected with proB cells isolated from primary
328 xenografts. Principal Component Analysis (PCA) showed clustering of samples based on
329 disease stage. Interestingly, diagnostic cells also clustered separately based on
330 differentiation state, while all relapse populations clustered closely together (Figure 5C). This
331 finding resonates well with our observation that in this patient leukaemia, relapse but not
332 diagnostic cells displayed high cell plasticity. While the PC1 component of the PC captured
333 most of the sample variance (47.15), PC2 clearly separated diagnostic proB cells and all
334 relapse subpopulations from diagnostic stemB cells. Pathways analysis on the genes driving
335 this PC component revealed enrichment for TNF-alpha via NF-KB signalling cascade, which
336 has previously been implicated in regulation of cell survival, proliferation and stemness[17].
337 Gene set enrichment analysis (GSEA) using publicly available gene expression
338 signatures[18,19] also validated our FACS analysis results in that compared to diagnostic proB
339 cells and relapse cells, diagnostic stemB cells downregulated cell proliferation associated
340 genes and expressed higher levels of markers typical of more immature stages of the
341 haematopoietic and B-cell differentiation hierarchy (such as HSC and MLP cells) (Figure 5C-
342 D and Supplementary Figure 3A-B). Most B-lineage cells from ALL patients are thought to
343 be arrested at some point of the transition from the proB to preB stage, therefore broadly
344 allowing the classification of ALL leukaemias into two distinct subtypes based on preBCR
345 function. Those arrested at the preBI stage rely upon IL7R/STAT5 signaling (preBCR$^-$), while
346 those arrested at preBII stage signal by the preBCR (preBCR$^+$)[20]. The frequency of pre-
347 BCR$^+$ or pre-BCR$^-$ cases has also been suggested to associate with different cytogenetic
348 subtypes of childhood ALL. For example, the majority of BCR-ABL1$^+$ leukaemias tend to
349 display constitutive active cytokine signalling via activation of STAT5 and repression of
350 BCL6. Herein, however, we found that these two genetically and phenotypically distinct
351 disease subsets can coexist within the same leukaemia. Interestingly, contrary to
352 expectations, diagnostic proB cells, which carried the BCR-ABL translocation and expressed
353 markers associated with later stages of the B-cell differentiation (Figure 5C), displayed

354 stronger expression of the preBCR receptor components and signalling cascade, while the
355 BCR-ABL$^-$ diagnostic stemB cells preferentially upregulated various genes involved in the
356 STAT5 signalling pathway (Supplementary Figure 3D).
357
358 We subsequently used SNP arrays to genetically characterise different phenotypic
359 compartments at presentation, relapse and after serial transplantation and track the
360 phylogenetic origins of the relapse propagating clones (Figure 6). SNP analysis of the
361 relapse samples indicated that it lacked several of the distinctive genetic abnormalities which
362 characterised either the dominant stemB compartment at diagnosis or the dominant proB.
363 However, the relapse hierarchy shared with the diagnostic compartments deletions on both
364 chromosomes 12 and 14. Of note, diagnosis and relapse clones also shared a common VDJ
365 rearrangement (data not shown).
366
367 Altogether, this data implies the survival over treatment of a common evolutionary ancestor
368 not detectable at diagnosis which gave rise to the relapse hierarchy through further
369 diversification. Accordingly, analysis of a bulk remission sample identified mosaic deletion of
370 chromosome 14 in the absence of del12 and any other obvious genetic abnormality.
371 Therefore, suggesting that this mosaic deletion is not a germline event and, likely, the
372 earliest genomic alteration (also preceding del12). To understand the basis of the mosaicism
373 observed in the remission sample, we separated CD3+ and CD3- cells and performed SNP
374 analysis. The CD3- compartment contained no evidence of the 14del. In contrast the CD3+
375 cells showed a complex pattern of chromosome 14 deletions indicative of a progressive
376 multi-step evolutionary process (Supplementary table 2). We next traced these 14del
377 variants back to the diagnostic and relapse stemB and proB compartments, as well as the
378 CD3+ compartment isolated at diagnosis. This spectrum of 14 deletions provides a branched
379 framework for subsequent evolution of stemB and proB clones seen at diagnosis and
380 relapse. Thus, using chromosomal deletions as markers of lineal ancestry we have here
381 reconstructed the evolutionary history of a Down's ALL from the predisposing event (Trisomy
382 21), through the acquisition of initiating lesions giving rise to a preleukemic hierarchy, into
383 further transformation marked by the onset of frank leukaemia. We also monitored the
384 response of this complex leukemic hierarchy to chemotherapy through the analysis of
385 matching remission and relapse specimens. Beyond the genetic considerations our work
386 emphasizes the role of cellular state primarily at the level of dormancy and cell cycle
387 indicating vulnerability to chemotherapy.
388
389 **Discussion**

390 Many studies have explored the extent and impact of different, individually assessed,
391 sources of intratumor heterogeneity on the evolution of tumours (particularly under
392 treatment). Our work, however, argues that cancer should be viewed as a complex matrix of
393 genetic and epigenetic heterogeneity such as cellular phenotype and cell cycle properties
394 which, altogether rather than in isolation, provide a substrate for tumour evolution during
395 disease progression in response to the selective impetus of therapeutic intervention.
396
397 The idea that epigenetically determined cell states, contribute to functional heterogeneity
398 amongst tumour cells by controlling fundamental cellular properties such as, for example,
399 cell identity, phenotype and cell-cycle stage is paradigmatically exemplified by the cancer
400 stem cell (CSC) hypothesis; which postulates that cells within a tumour are organized in a
401 hierarchical fashion reflecting lineage relationships and tumorigenic potential, and that the
402 maintenance of cancer clones is uniquely dependent on the most primitive CSCs with self-
403 renewal capacity[21–23]. This notion, which was pioneered in the context of blood malignancies,
404 remains however extensively debated. While phenotypic heterogeneity - primarily at the level
405 of immune-phenotype – has been widely observed in ALL, disparate data exist concerning
406 the existence of LSCs within this disease. Earlier work by Rehe et al., showed that many
407 cellular fractions contain LSC-potential at a similar frequency and that the majority of - if not
408 all - cells may be able to propagate the disease[24]. Importantly, the same study also found
409 evidence for phenotypic plasticity within different subpopulations upon serial transplantation
410 in immune-deficient mice. Distinct leukaemia stem cells (LSCs) have nonetheless been
411 identified in acute myeloid leukaemia (AML), where functional evidence for their existence
412 has been provided[25–28]. However, even in this context, their exact phenotype remains
413 elusive, with more recent studies uncovering LSC activity across different cellular subsets[29].
414 We here provide evidence that although different Down's ALL disease compartments can all
415 contribute to leukaemia propagation *in vivo,* these same compartments display largely
416 distinct fitness; both in the presence and in the absence of chemotherapy. In this scenario,
417 more primitive stem-like compartments are outcompeted by highly proliferative early-B cell
418 progenitors in the absence of treatment. However, this selective advantage is reversed in the
419 presence of chemotherapy, which selects for quiescent cells. While the genetically
420 heterogenous nature of childhood ALL has been demonstrated by multiple earlier studies,
421 our current work provides the first direct evidence for genetic diversity of cancer propagating
422 cells in patients with Down's syndrome ALL. Our SNP array analysis of serially transplanted
423 ALLs shows that leukemic cells undergo a process of gradual branching clonal evolution
424 demonstrated by the coexistence within a leukaemia of multiple subclones bearing
425 alterations of the same genomic locus characterised by distinct breakpoints.

426

427 Our earlier work showed that in many ALL leukemias genotype and phenotypes with
428 relevance to chemoresistance do not segregate. As a result, induction chemotherapy does
429 not select at the level of genotype. We here identified and studied the evolutionary history of
430 an unusual case of BCR-ABL+ Down's ALL in which genotype and phenotype do instead
431 clearly segregate. In this patient's leukaemia the stem/B cells were fusion-gene-negative
432 while proB/pre-B cells were fusion-gene-positive. This case allowed us to reconstruct in
433 detail the complex genotype-phenotype relationship characterizing different leukemic cell
434 fractions labelled by both distinct genetic makeup and fitness. We hence observed that
435 highly quiescent and developmentally more primitive cells preferentially escaped treatment.
436 As these cells carried only a subset of diagnostic genetic alterations, the relapse disease
437 was genetically divergent from the diagnostic specimen. Reconstruction of the leukaemia
438 phylogenetic tree demonstrated that the dominant population at relapse originated from a
439 rare pre-leukemic clone. This was likely present at diagnosis below detection level and, in
440 line with its likely low proliferative activity, did not extensively expand upon transplantation.
441

442 We first demonstrated the existence of pre-leukemic cells in ALL in studies of mono-
443 chorionic twins of whom one was diagnosed with TEL-AML1-positive ALL[15,30]. Blood spot
444 studies from Guthrie-cards further showed that these pre-leukemic cells can precede overt
445 leukaemia by up to 12 years and are also detectable in patients that never go on to overt
446 disease[31,32]. A case report of a 7-year old boy diagnosed with ALL with a mixed lympho-
447 myeloid phenotype previously suggested that BCR-ABL+ non-Down leukaemic cells could
448 reactivate to drive disease recurrence more than two decades from initial diagnosis[33]. Our
449 new data add to our understanding of this elusive population by showing that Down
450 Syndrome ALL chr21+ cells, which have not yet acquired the BCR-ABL translocation
451 labelling the more proliferative proB compartment disease, can preferentially survive
452 chemotherapy and, therefore, represent a reservoir for relapse initiation. Whether these cells
453 are quiescent principally as a function of their unique developmental origin[34] or rather of their
454 genotype[35] or a combination of both remains to be investigated and would be important to
455 understand if new therapeutic interventions targeting these cells are to be designed. This
456 notwithstanding, pre-leukemic clones, even when rare, should clearly be considered part of
457 the fully transformed leukaemia landscape and may upon selection by therapy provide an
458 additional cellular source for subsequent relapse.
459

460 **Acknowledgments**

We are grateful for the support of Leukaemia and Lymphoma Research (TE; Specialist Programme Grant), Cancerfonden (TE; Project grants CAN 2013/712 and 20 1334 PjF) , MRC Molecular Haematology Unit core award (SEJ; MC_UU_12009/5), Blood Cancer UK (SEJ), Swedish Childhood Cancer Fund (SEJ), the National Institute for Health Research Biomedical Research Centre Programme, the CBRC, UCL, Haematolinne, the Leukemia and Lymphoma Society, USA (Career Development Program fellowship to PSW), the Deutsche Forschungsgemeinschaft (fellowship to CL; LU1474/1-1 and programme grant as part of the Sonderforschungsbereich; SFB873, Teilprojekt A13).

Authorship Contributions

CL and VT performed experiments; RC analysed the SNP array; JH analysed the RNAseq data with input from VT; AS interpreted and supervised SNP-array data analyses; TS developed the mathematical model; CL and PSW performed the cell cycle analysis, VB carried out FISH analyses; AC provided patient data and samples; SAC, HF and PSW performed cell sorting. CL, VT, SEWJ and TE designed the study and interpreted data. CL, VT and TE wrote the paper with critical input from other authors.

Conflict of Interest Disclosures

The authors declare no conflict of interest.

References

1. Rosenthal, R., Mcgranahan, N., Herrero, J. & Swanton, C. Deciphering Genetic Intratumor Heterogeneity and Its Impact on Cancer Evolution. *Annu. Rev. Cancer Biol. 2017* **1**, 223–240 (2016).
2. Marusyk, A., Almendro, V. & Polyak, K. Intra-tumour heterogeneity: A looking glass for cancer? *Nature Reviews Cancer* vol. 12 323–334 (2012).
3. Landau, D. A., Carter, S. L., Getz, G. & Wu, C. J. Clonal evolution in hematological malignancies and therapeutic implications. *Leukemia* vol. 28 34–43 (2014).
4. Anderson, K. *et al.* Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* **469**, 356–361 (2011).
5. Notta, F. *et al.* Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells. *Nature* **469**, 362–7 (2011).
6. Turati, V. A. *et al.* Chemotherapy induces canalization of cell state in childhood B-cell precursor acute lymphoblastic leukemia. *Nat. Cancer* doi:10.1038/s43018-021-00219-3.
7. Ebinger, S. *et al.* Characterization of Rare, Dormant, and Therapy-Resistant Cells in

496 Acute Lymphoblastic Leukemia. *Cancer Cell* **30**, 849–862 (2016).
497 8. Lutz, C. *et al.* Quiescent leukaemic cells account for minimal residual disease in
498 childhood lymphoblastic leukaemia. *Leukemia* (2013) doi:10.1038/leu.2012.306.
499 9. Waanders, E. *et al.* Mutational Landscape and Patterns of Clonal Evolution in
500 Relapsed Pediatric Acute Lymphoblastic Leukemia. *Blood Cancer Discov.* **1**, 96–111
501 (2020).
502 10. Dobson, S. M. *et al.* Relapse-Fated Latent Diagnosis Subclones in Acute B Lineage
503 Leukemia Are Drug Tolerant and Possess Distinct Metabolic Programs. *Cancer*
504 *Discov.* (2020) doi:10.1158/2159-8290.cd-19-1059.
505 11. Kirschner-Schwabe, R. *et al.* Chemotherapy and mismatch repair deficiency
506 cooperate to fuel TP53 mutagenesis and ALL relapse. *Nat. Cancer*
507 doi:10.1038/s43018-021-00230-8.
508 12. Buitenkamp, T. D. *et al.* Acute lymphoblastic leukemia in children with Down
509 syndrome: A retrospective analysis from the Ponte di Legno study group. *Blood* **123**,
510 70–77 (2014).
511 13. Xavier, A. C. & Taub, J. W. Acute leukemia in children with down syndrome.
512 *Haematologica* vol. 95 1043–1045 (2010).
513 14. Schuh, A. *et al.* Monitoring chronic lymphocytic leukemia progression by whole
514 genome sequencing reveals heterogeneous clonal evolution patterns. *Blood* **120**,
515 4191–4196 (2012).
516 15. Hong, D. *et al.* Initiating and cancer-propogating cells in TEL-AML1-associated
517 childhood leukemia. *Science (80-.).* **319**, 336–339 (2008).
518 16. Zur Wiesch, P. S., Engelstädter, J. & Bonhoeffer, S. Compensation of fitness costs
519 and reversibility of antibiotic resistance mutations. *Antimicrob. Agents Chemother.* **54**,
520 2085–2095 (2010).
521 17. Vazquez-Santillan, K., Melendez-Zajgla, & J., Jimenez-Hernandez, & L., Martínez-
522 Ruiz, G. & Maldonado, & V. NF-κB signaling in cancer stem cells: a promising
523 therapeutic target? doi:10.1007/s13402-015-0236-6.
524 18. Sergushichev, A. A. An algorithm for fast preranked gene set enrichment analysis
525 using cumulative statistic calculation. *bioRxiv* 060012 (2016).
526 19. Laurenti, E. *et al.* The transcriptional architecture of early human hematopoiesis
527 identifies multilevel control of lymphoid commitment. *Nat. Immunol.* **14**, 756–763
528 (2013).
529 20. Geng, H. *et al.* Self-Enforcing Feedback Activation between BCL6 and Pre-B Cell
530 Receptor Signaling Defines a Distinct Subtype of Acute Lymphoblastic Leukemia.
531 *Cancer Cell* **27**, 409–425 (2015).

532 21. Dick, J. E. Stem cell concepts renew cancer research. *Blood* **112**, 4793–4807 (2008).
533 22. Kreso, A. & Dick, J. E. Evolution of the cancer stem cell model. *Cell Stem Cell* **14**,
534 275–291 (2014).
535 23. Ailles, L. E. & Weissman, I. L. Cancer stem cells in solid tumors. *Curr. Opin.*
536 *Biotechnol.* **18**, 460–466 (2007).
537 24. Rehe, K. *et al.* Acute B lymphoblastic leukaemia-propagating cells are present at high
538 frequency in diverse lymphoblast populations. *EMBO Mol. Med.* **5**, 38–51 (2013).
539 25. Shlush, L. I. *et al.* Tracing the origins of relapse in acute myeloid leukaemia to stem
540 cells. *Nature* **547**, 104–108 (2017).
541 26. Velten, L. *et al.* Identification of leukemic and pre-leukemic stem cells by clonal
542 tracking from single-cell transcriptomics. *Nat. Commun.* **12**, 1–13 (2021).
543 27. Lapidot, T. *et al.* A cell initiating human acute myeloid leukaemia after transplantation
544 into SCID mice. *MacKinnon, R. Yallen. G. Sci.* **362**, 31–39 (1993).
545 28. Dick, J. E. & Bonnet, D. Human Acute Myeloid Leukaemia is organised as a heirarchy
546 that originates from a primitive haematopoetic cell. *Nat. Med.* **3**, 730–737 (1997).
547 29. Taussig, D. C. *et al.* Leukemia-initiating cells from some acute myeloid leukemia
548 patients with mutated nucleophosmin reside in the CD34- fraction. *Blood* **115**, 1976–
549 1984 (2010).
550 30. Ford, A. M. *et al.* Fetal origins of the TEL-AML1 fusion gene in identical twins with
551 leukemia. *Med. Sci.* **95**, 4584–4588 (1998).
552 31. Greaves, M. F., Maia, A. T., Wiemels, J. L. & Ford, A. M. Leukemia in twins: Lessons
553 in natural history. *Blood* **102**, 2321–2333 (2003).
554 32. Gale, K. B. *et al.* Backtracking leukemia to birth: Identification of clonotypic gene
555 fusion sequences in neonatal blood spots. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 13950–
556 13954 (1997).
557 33. Ford, A. M. *et al.* Protracted dormancy of pre-leukemic stem cells. *Leukemia* **29**,
558 2202–2207 (2015).
559 34. Böiers, C. *et al.* A Human IPS Model Implicates Embryonic B-Myeloid Fate Restriction
560 as Developmental Susceptibility to B Acute Lymphoblastic Leukemia-Associated
561 ETV6-RUNX1. *Dev. Cell* **44**, 362-377.e7 (2018).
562 35. Wray, J. P. et al. Cell cycle corruption in a pre-leukemic ETV6-RUNX1 model exposes
563 RUNX1 addiction as a therapeutic target in acute lymphoblastic leukemia. *bioRxiv*
564 *Prepr.* (2021).
565 36. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human
566 oligodendroglioma. *Nat. Publ. Gr.* **539**, (2016).
567

568

569 **Figure legends**

570 **Figure 1: Study of clonal dynamics through serial transplantation assays. A.**
571 Experimental setup. BM samples before and after serial transplantation in NSG-mice were
572 analyzed by SNP-arrays. **B.** SNP array results of patient 1. Deletions on Chromosomes 1,
573 12 and a single amplification on Chromosome 17 were identified and shown to be stable
574 during serial transplantation. Relationships between transplants recipients are indicated with
575 arrows. **C.** SNP array results of patient 2. Deletions on Chromosomes 9, 14 and a single
576 amplification on Chromosome X were identified. During serial transplantation copy number
577 abnormalities (CNA) disappeared and re-appeared suggesting the co-existence of different
578 sub-clones. Relationships between transplants recipients are indicated with arrows. **D.** Initial
579 reconstruction of the clonal hierarchy of patient 2 leukemia considering Chromosomes 9, 14
580 and X. * Indicate secondary transplants from a primary transplant for which SNP-array data
581 are unavailable due to insufficient numbers of cells for analysis.

582

583 **Figure 2: Detailed analysis on chromosome 9 reveals higher clonal complexity**
584 **(Patient 2). A.** In depth analysis of CNA on chromosome 9 during serial transplantation
585 suggests a more complex clonal hierarchy with multiple deletions of different size targeting
586 the locus and occurrence of cnLOH. Each box indicates an individual mouse. **B.**
587 Phylogenetic tree of leukemia cells based on SNP-array data obtained at diagnosis and
588 during serial transplantation. Mosaicim of CNA suggested the existence of sub-clones within
589 analyzed samples. The presented tree represents the minimum number of diagnostic clones
590 required to explain the mutational landscape identified in the engrafted leukaemias. Minimal
591 clone numbers were calculated by mathematical modelling (see supplementary materials).

592

593 **Figure 3: Correlation of phenotype and genotype. A.** Experimental setup. BM bulk,
594 StemB ($CD34^{+}CD38^{-/low}CD19^{+}$) and ProB ($CD34^{+}CD38^{+}CD19^{+}$) cells were analyzed by SNP-
595 arrays before and (if available) after serial transplantation in NSG mice. **B.** In patient 2 bulk,
596 stemB and ProB cells showed identical CNA. **C.** In Patient 3 animals transplanted with bulk,
597 stemB, and ProB cells showed identical CNA profiles: suggesting the existence of a single
598 dominant clone. * Indicate secondary transplants from a primary transplant for which SNP-
599 array data are unavailable due to insufficient numbers of cells for analysis. Secondary
600 transplants from Stem/B and bulk cells did not engraft.

601

602 **Figure 4: Phenotypically, genetically and functionally distinct sub-clones co-exist in a**
603 **case of BCR-ABL1^{+} ALL. A.** FISH analysis of leukemic subpopulations at diagnosis and

604 after serial xenotransplantation in Patient 4 using a BCR/ABL-specific probe. To confirm the
605 leukemic status, sorted cells were analysed by FISH for the appropriate clonal marker.
606 Presence of the t(9;22) translocation is indicated by detection of one green, one red and two
607 red-green fusion signals. **B.** Competitive transplants of genetically and phenotypically distinct
608 BCR-ABL$^+$ ProB (CD34$^+$CD38$^+$CD19$^+$) and BCR-ABL$^-$ StemB (CD34$^+$CD38$^{-/low}$CD19$^+$) cells.
609 Input and output ratios of the two populations in individual mice are displayed. **C.** Cell cycle
610 analysis of leukemic ProB (upper panels) and StemB (lower panels) cells from competitively
611 transplanted NSG-mice with subpopulations from a BCR-ABL1$^+$ ALL patient. Percentages of
612 cells in G_0 (DAPI$^-$Ki67$^-$), G_1 (DAPI$^-$Ki67$^+$) and S/G_2/M (DAPI$^+$Ki67$^+$) are shown.
613
614 **Figure 5: Selective therapy resistance of CD34$^+$CD38$^{-/low}$CD19$^+$ Stem/B cells in high**
615 **risk BCR-ABL1$^+$ ALL. A.** FACS analysis of BCR-ABL1 ALL subpopulations at diagnosis,
616 remission, and relapse (patient 4). Percentages of CD34$^+$CD38$^{-/low}$CD19$^-$ (HSC),
617 CD34$^+$CD38$^{-/low}$CD19$^+$ (StemB), CD34$^+$CD38$^+$CD19$^+$ (ProB) cells and CD34$^+$CD38$^+$CD19$^-$
618 (Progenitors) in total bone marrow (BM) mononuclear cells (MNC) are shown. The black
619 boxes show the gates used for frequency calculations. Not all collected events are shown. **B.**
620 FACS analysis of BM cells from serially transplanted NSG mice injected with sorted StemB
621 and ProB cells from patient 4 relapse. **C.** PCA analysis based on bulk RNA sequencing of
622 diagnostic and relapse cells harvested from primary and secondary xenografts. Six different
623 populations of cells were analysed. i) diagnostic stemB cells, ii) diagnostic proB cells iii)
624 relapse stemB cells harvested from secondary xenografts injected with stemB cells isolated
625 from primary xenografts iv) relapse proB cells harvested from secondary xenografts injected
626 with stemB cells isolated from primary xenografts, v) relapse proB cells harvested from
627 secondary xenografts injected with proB cells isolated from primary xenografts, and vi)
628 relapse stemB cells harvested from secondary xenografts injected with proB cells isolated
629 from primary xenografts. Arrows showing the contribution of different gene expression
630 signatures (calculated as per panel D) to the sample's clusters biology are shown. **D.** Gene
631 set enrichment analysis (GSEA) for a selection of MSigDB-50 Hallmark gene-sets and
632 published signatures associated with predefined differentiation stages of the B-cell
633 ontogeny[19]. ProB diagnostic cells are compared to the stemB diagnostic cells. Significance
634 thresholds: <0.05 after adjusting for multiple testing (using the Benjamini-Hochberg method).
635
636 **Figure 6: Distinct sub-clones in the BCR-ABL1$^+$ ALL and the BCR-ABL$^-$ relapse clones**
637 **are genetically related (Patient 4). A.** SNP array results of diagnostic stem/B, xenografted
638 stem/B, diagnostic proB, xenografted proB, diagnostic preB, relapse StemB, relapse ProB
639 and MRD-negative Remission sample are displayed. A chromosome 12 deletion was

640 present in all but the remission control sample illustrating the clonal relationship of all
641 leukemic samples. CNA of diagnostic stem/B and ProB were identified and stable during
642 serial transplantation with identification of some mosaicism suggesting sub-clones in each
643 compartment. Relationships of transplants are indicated. Relapse stem/B and proB were
644 genetically identical but stem/B showed 90% mosaicism of all clones suggesting the
645 existence of a pre-leukemic clone. **B.** Detailed analyses SNP array revealed different sizes
646 of some chromosome deletions at diagnosis, remission and relapse and suggest the
647 existence of distinct pre-leukemic sub-clones for which a clonal hierarchy cand be re-
648 constructed.
649
650 **Figure 7: Pre-leukemic and leukemic hierarchy.** Based on genetic mosaicism and
651 different sizes of chromosome 7 and 14 deletions (Patient 4) a pre-leukemic clonal hierarchy
652 could be constructed which showed clonal evolution at the pre-leukemic stage that gave rise
653 to the diagnostic leukemic sub-clones as well as evolving into another pre-leukemic sub-
654 clone responsible for relapse.
655
656 **Supplementary Figures Legends**
657 **Figure S1: Flow cytometry sorting strategy.** Representative gating strategy used for the
658 analysis of immunophenotypically-distinct Down's syndrome ALL subpopulations before and
659 after xenotransplantation. StemB ($CD34^{+}CD38^{-/low}CD19^{+}$) and proB ($CD34^{+}CD38^{+}CD19^{+}$)
660 cells were sorted. **A-B**. Patient 2 and 3 leukaemia respectively.
661 **Figure S2: Cell cycle analysis. A.** Cell cycle analysis of representative xenografted stemB
662 and ProB as compared to the **B.** patient relapse.
663 **Figure S3: RNA sequencing analysis. A-C** Gene Set Variation Analysis (GSVA) analysis
664 of RNA sequencing data looking at the expression of cell cycle related gene signatures. Six
665 different populations of cells were analysed: i) diagnostic stemB cells, ii) diagnostic proB
666 cells iii) relapse stemB cells harvested from secondary xenografts injected with stemB cells
667 isolated from primary xenografts iv) relapse proB cells harvested from secondary xenografts
668 injected with stemB cells isolated from primary xenografts, v) relapse proB cells harvested
669 from secondary xenografts injected with proB cells isolated from primary xenografts, and vi)
670 relapse stemB cells harvested from secondary xenografts injected with proB cells isolated
671 from primary xenografts. Within each heatmap samples are sorted by category and genes
672 split into top 500 most variable vs all others. **A.** G1/S gene signature[36], **B.** G2/M gene
673 signature[36], **C.** E2F targets gene signature. **C.** Gene set enrichment analysis (GSEA) for a
674 selection of MSigDB-50 Hallmark gene-sets and published signatures associated with
675 predefined differentiation stages of the B-cell ontogeny[19]. This analysis looks specifically at

676 the top 500 genes driving clustering along the PC2 component of the PCA analysis. **D.** PCA
677 analysis based on bulk RNA sequencing of diagnostic and relapse cells harvested from
678 primary and secondary xenografts. Arrows showing the contribution of gene expression
679 signatures involved in BCR signalling and B-cell differentiation are shown. Significance
680 thresholds: <0.05 after adjusting for multiple testing (using the Benjamini-Hochberg method).

Figure 1

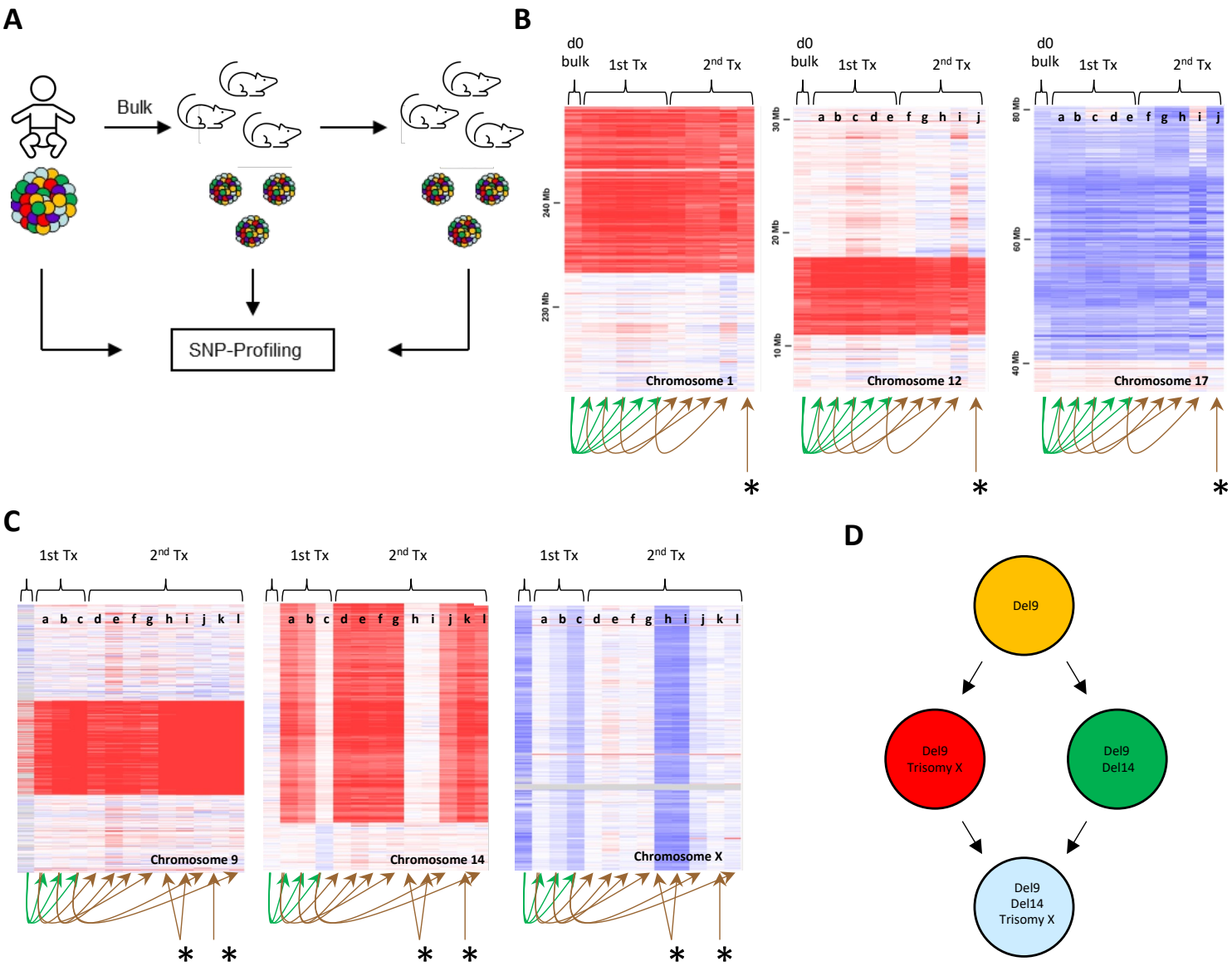
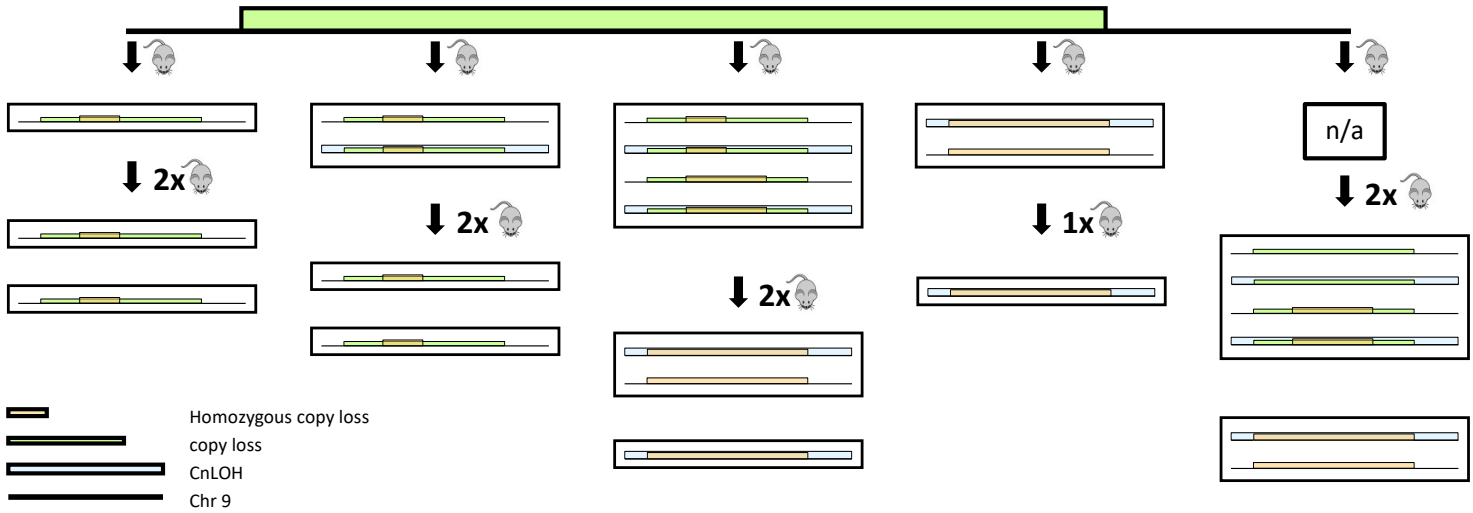

A
Bulk
SNP-Profiling
B
d0 bulk
1st Tx
2nd Tx
240 Mb
230 Mb
Chromosome 1
30 Mb
20 Mb
10 Mb
Chromosome 12
80 Mb
60 Mb
40 Mb
Chromosome 17
a b c d e f g h i j
C
a b c d e f g h i j k l
Chromosome 9
Chromosome 14
Chromosome X
D
Del9
Del9 Trisomy X
Del9 Del14
Del9 Del14 Trisomy X

Figure 2

A


2x
2x
2x
1x
2x
n/a
Homozygous copy loss
copy loss
CnLOH
Chr 9

B

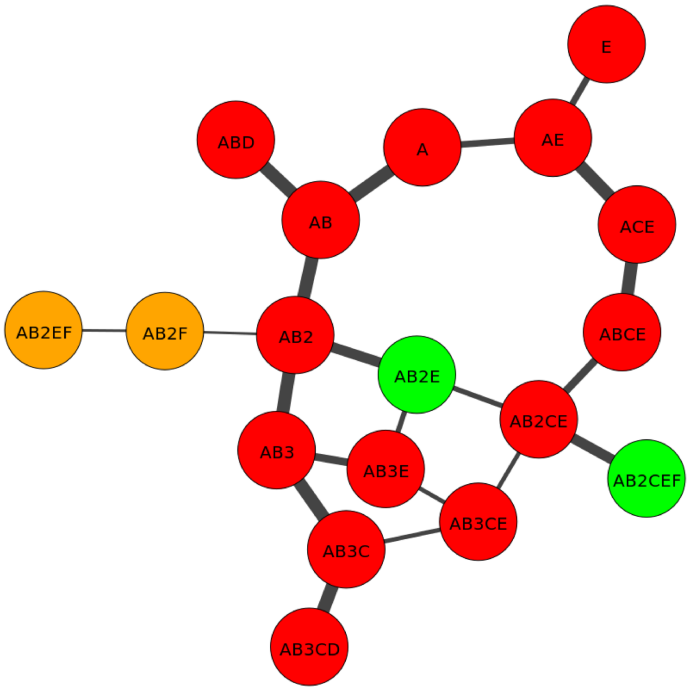

Connection present in
100%
80%
60%
40%
20%
of all graphs
Node present in
100%
80%
20%
of all graphs
E
ABD
A
AE
AB
ACE
AB2EF
AB2F
AB2
ABCE
AB2E
AB2CE
AB3
AB3E
AB2CEF
AB3CE
AB3C
AB3CD

Figure 3

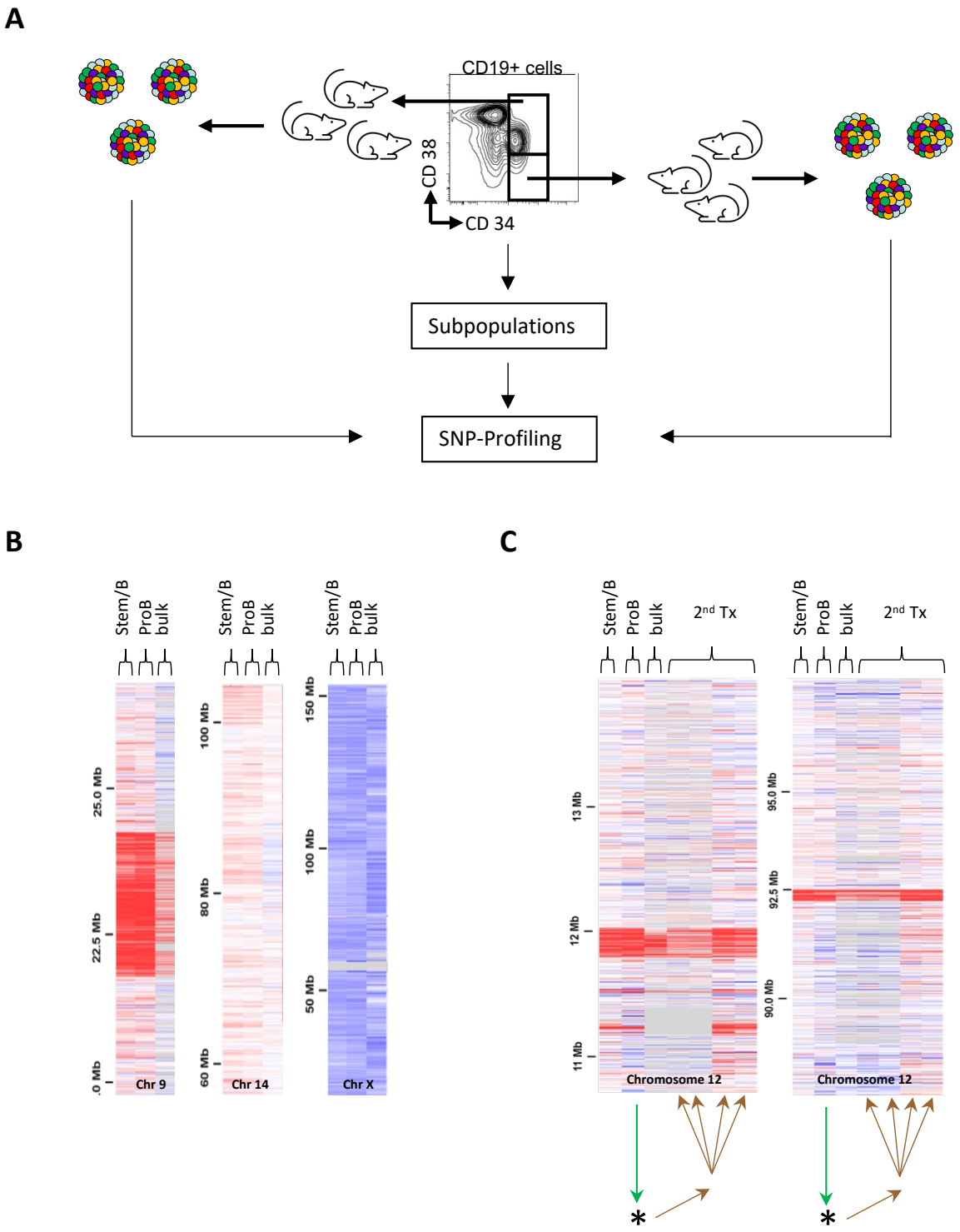

A
CD19+ cells
CD 38
CD 34
Subpopulations
SNP-Profiling
B
Stem/B
ProB
bulk
25.0 Mb
22.5 Mb
Chr 9
100 Mb
80 Mb
60 Mb
Chr 14
150 Mb
100 Mb
50 Mb
Chr X
C
2nd Tx
13 Mb
12 Mb
11 Mb
Chromosome 12
95.0 Mb
92.5 Mb
90.0 Mb
Chromosome 12
*
*

Figure 4

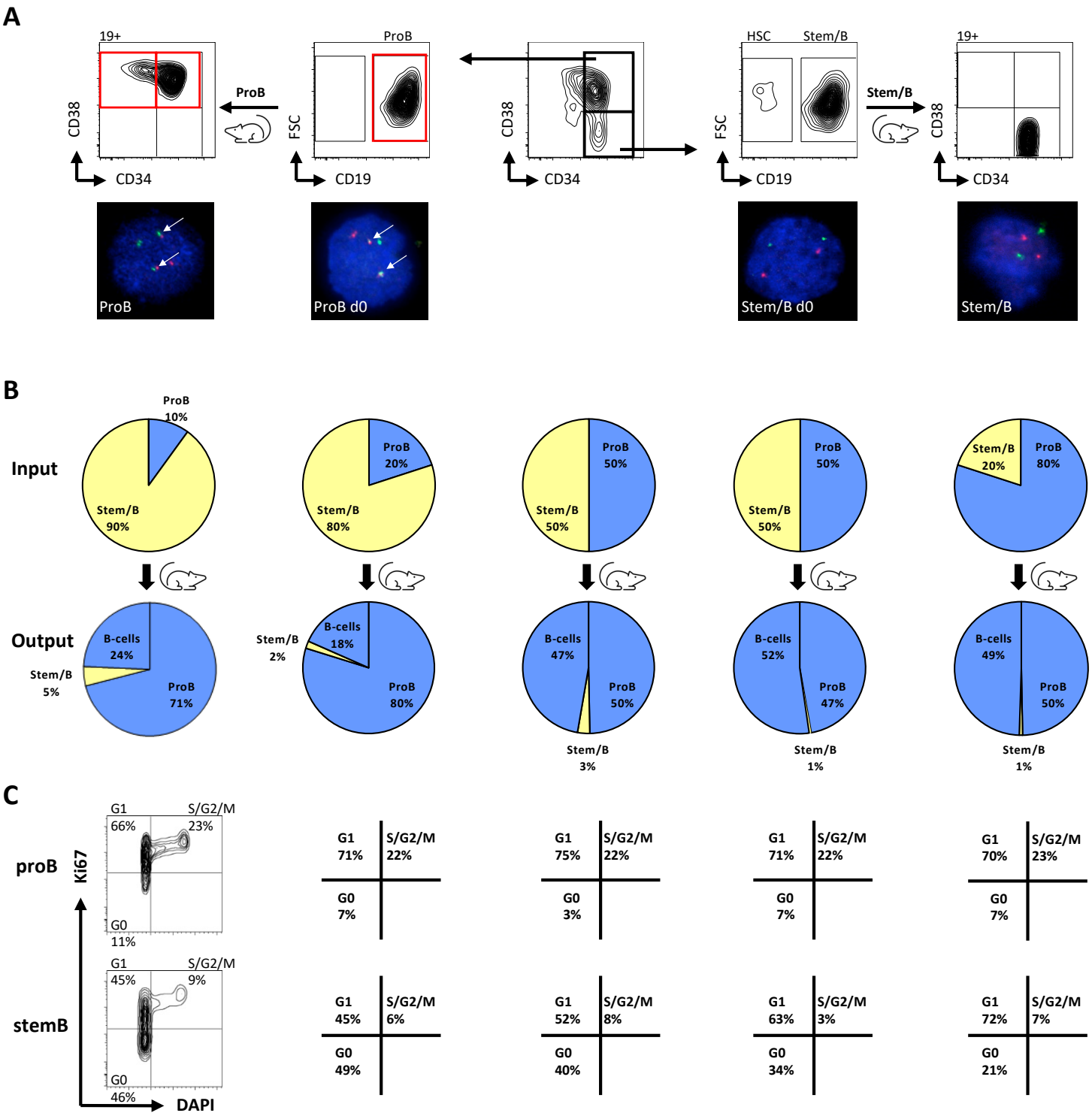

A
19+
CD38
CD34
ProB
ProB
FSC
CD19
CD38
CD34
HSC
Stem/B
FSC
CD19
Stem/B
19+
CD38
CD34
ProB
ProB d0
Stem/B d0
Stem/B
B
Input
Output
ProB 10%
Stem/B 90%
ProB 20%
Stem/B 80%
ProB 50%
Stem/B 50%
ProB 50%
Stem/B 50%
Stem/B 20%
ProB 80%
B-cells 24%
Stem/B 5%
ProB 71%
Stem/B 2%
B-cells 18%
ProB 80%
B-cells 47%
ProB 50%
Stem/B 3%
B-cells 52%
ProB 47%
Stem/B 1%
B-cells 49%
ProB 50%
Stem/B 1%
C
proB
stemB
Ki67
DAPI
G1 66%
S/G2/M 23%
G0 11%
G1 45%
S/G2/M 9%
G0 46%
G1 71%
S/G2/M 22%
G0 7%
G1 45%
S/G2/M 6%
G0 49%
G1 75%
S/G2/M 22%
G0 3%
G1 52%
S/G2/M 8%
G0 40%
G1 71%
S/G2/M 22%
G0 7%
G1 63%
S/G2/M 3%
G0 34%
G1 70%
S/G2/M 23%
G0 7%
G1 72%
S/G2/M 7%
G0 21%

Figure 5

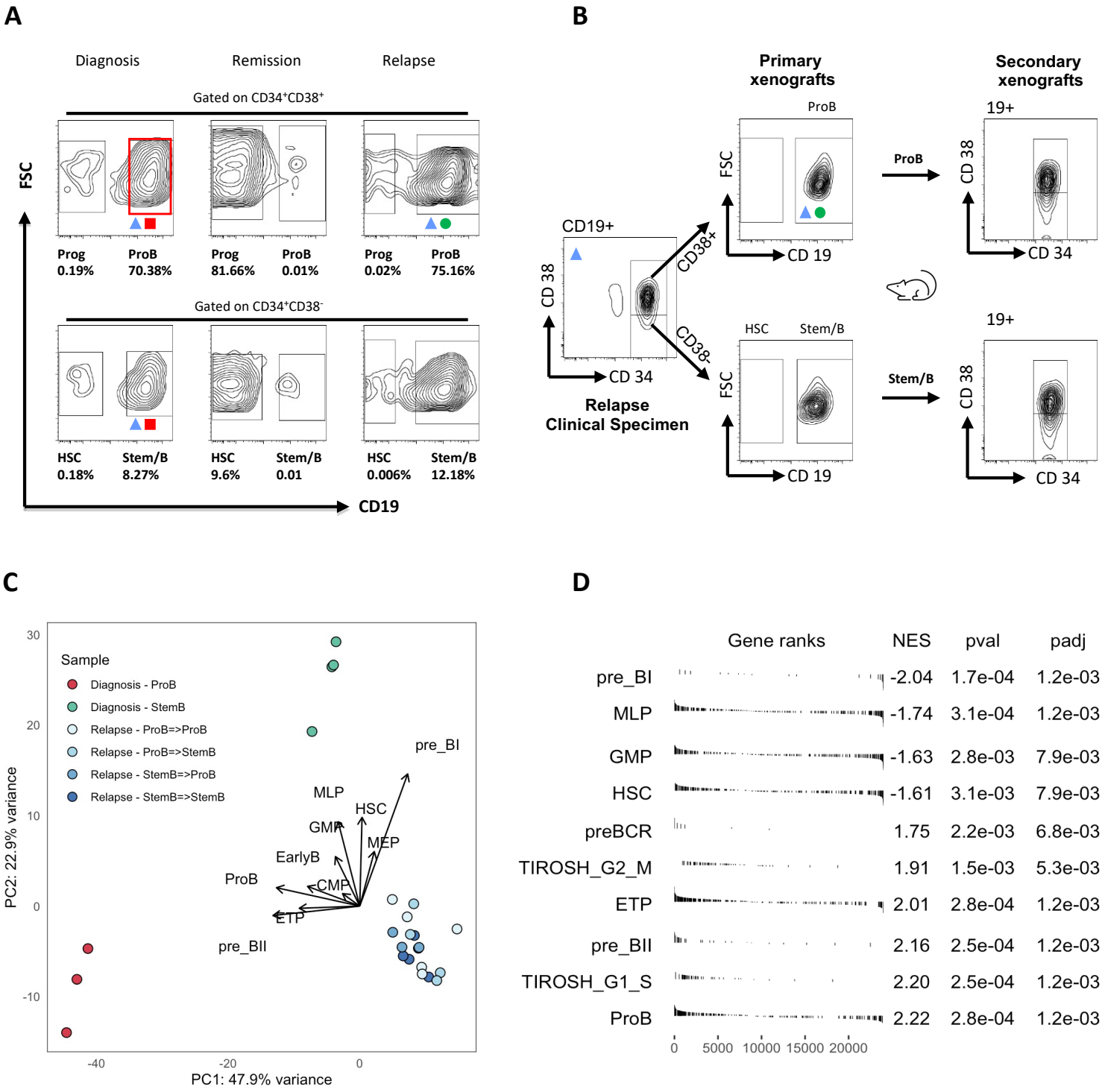

A
Diagnosis
Remission
Relapse
Gated on CD34+CD38+
FSC
Prog 0.19%
ProB 70.38%
Prog 81.66%
ProB 0.01%
Prog 0.02%
ProB 75.16%
Gated on CD34+CD38-
HSC 0.18%
Stem/B 8.27%
HSC 9.6%
Stem/B 0.01
HSC 0.006%
Stem/B 12.18%
CD19
B
Primary xenografts
Secondary xenografts
ProB
19+
CD19+
CD38+
CD38-
CD 38
CD 34
FSC
CD 19
HSC
Stem/B
Relapse Clinical Specimen
C
Sample
Diagnosis - ProB
Diagnosis - StemB
Relapse - ProB=>ProB
Relapse - ProB=>StemB
Relapse - StemB=>ProB
Relapse - StemB=>StemB
pre_BI
MLP
HSC
GMP
MEP
EarlyB
ProB
CMP
ETP
pre_BII
PC2: 22.9% variance
PC1: 47.9% variance
30
20
10
0
-10
-40
-20
0
D
Gene ranks
NES
pval
padj
pre_BI -2.04 1.7e-04 1.2e-03
MLP -1.74 3.1e-04 1.2e-03
GMP -1.63 2.8e-03 7.9e-03
HSC -1.61 3.1e-03 7.9e-03
preBCR 1.75 2.2e-03 6.8e-03
TIROSH_G2_M 1.91 1.5e-03 5.3e-03
ETP 2.01 2.8e-04 1.2e-03
pre_BII 2.16 2.5e-04 1.2e-03
TIROSH_G1_S 2.20 2.5e-04 1.2e-03
ProB 2.22 2.8e-04 1.2e-03
0
5000
10000
15000
20000

Figure 6

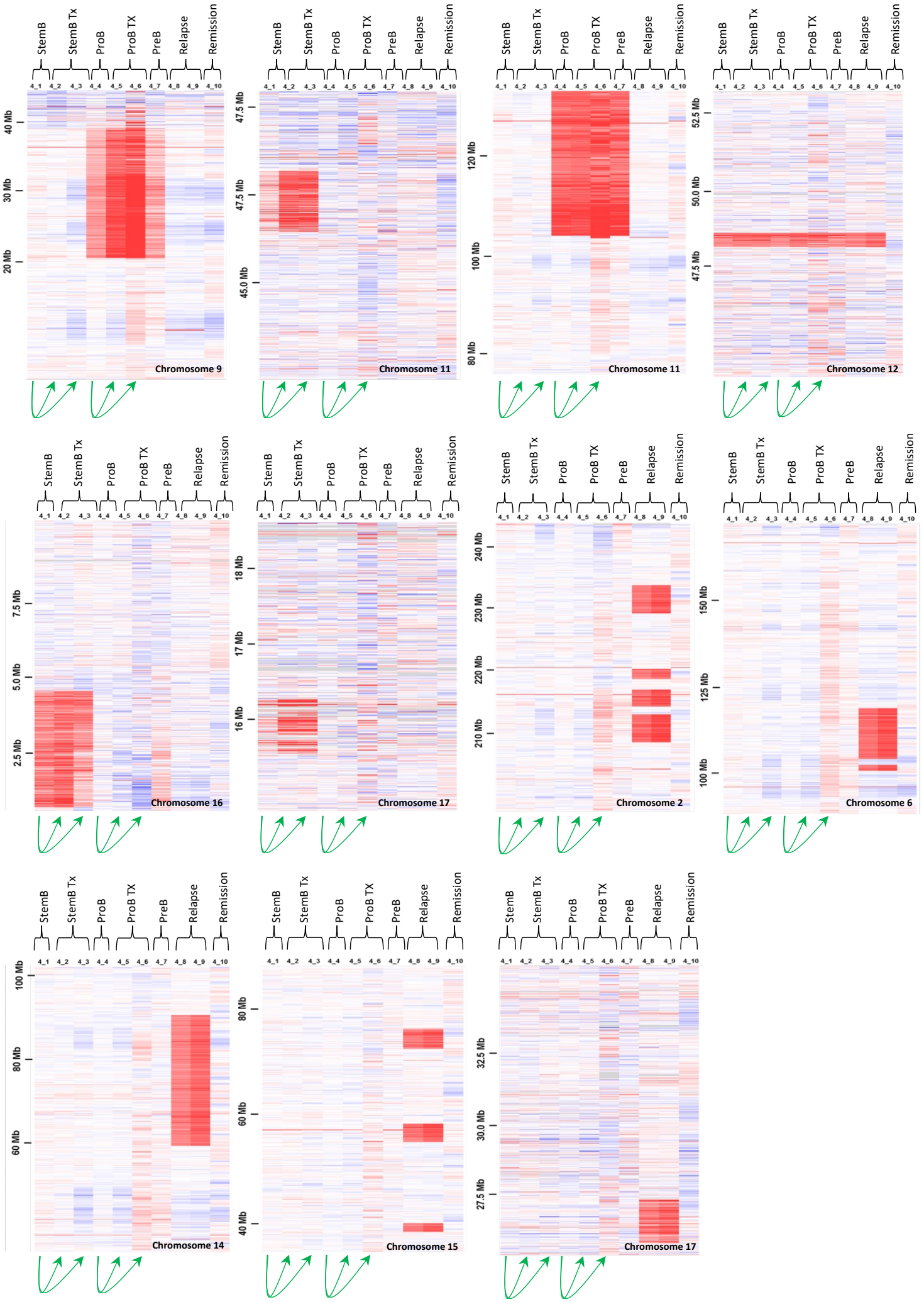
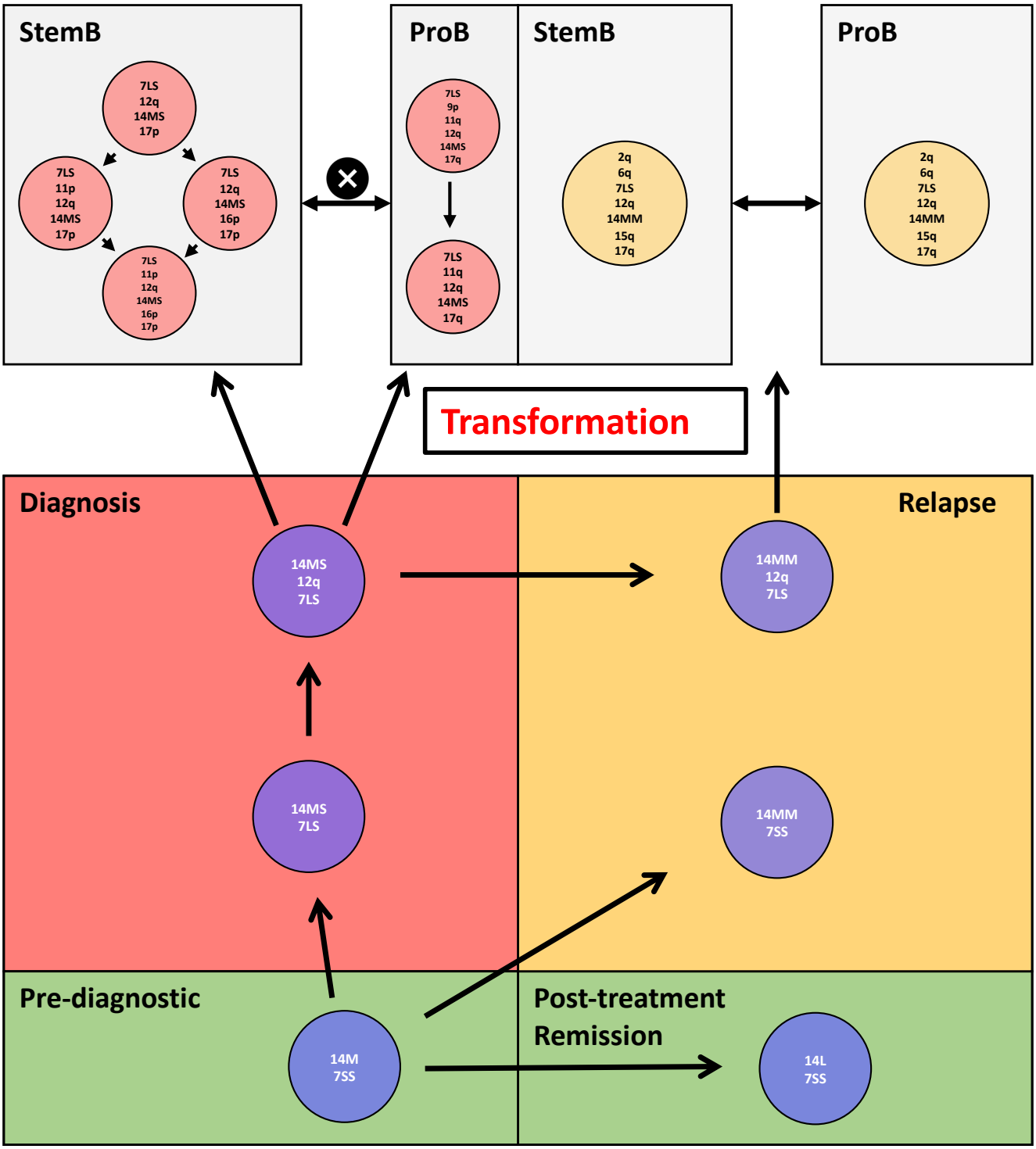
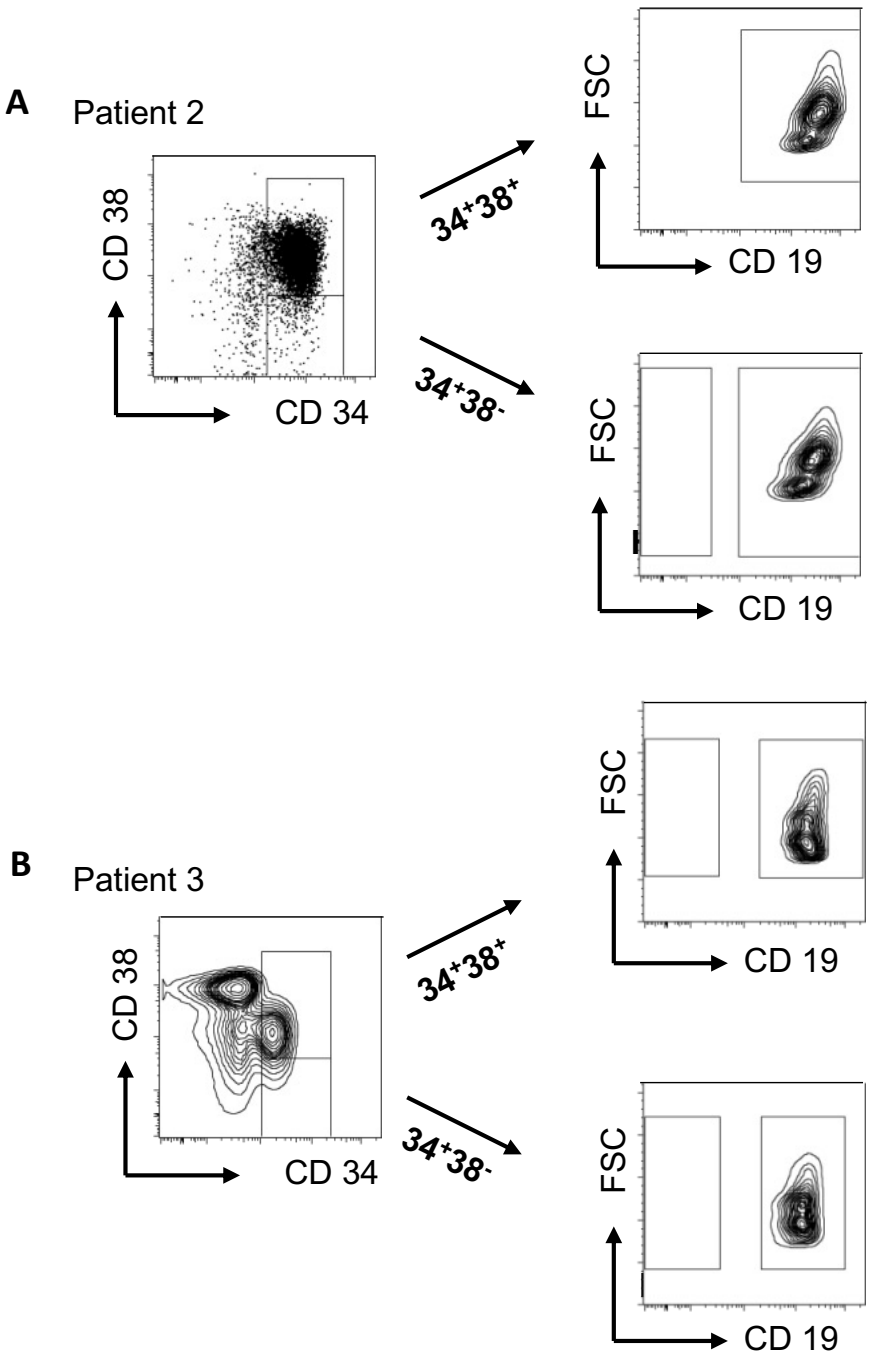

StemB
StemB Tx
ProB
ProB TX
PreB
Relapse
Remission
4_1 4_2 4_3 4_4 4_5 4_6 4_7 4_8 4_9 4_10
Chromosome 9
Chromosome 11
Chromosome 11
Chromosome 12
Chromosome 16
Chromosome 17
Chromosome 2
Chromosome 6
Chromosome 14
Chromosome 15
Chromosome 17

Figure 7

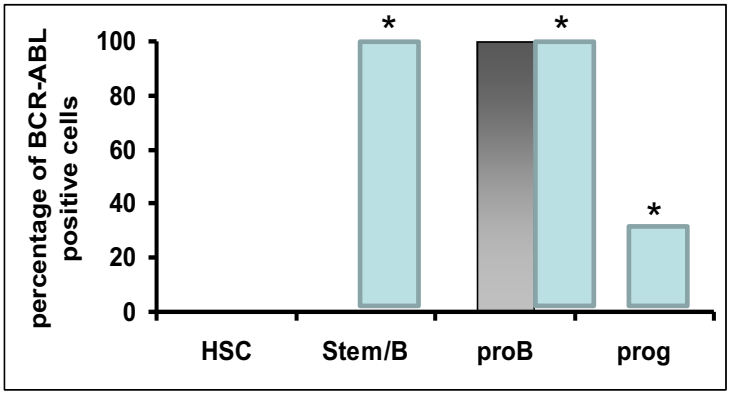

StemB
7LS
12q
14MS
17p
7LS
11p
12q
14MS
17p
7LS
12q
14MS
16p
17p
7LS
11p
12q
14MS
16p
17p
ProB
7LS
9p
11q
12q
14MS
17q
7LS
11q
12q
14MS
17q
StemB
2q
6q
7LS
12q
14MM
15q
17q
ProB
2q
6q
7LS
12q
14MM
15q
17q
Transformation
Diagnosis
Relapse
14MS
12q
7LS
14MM
12q
7LS
14MS
7LS
14MM
7SS
Pre-diagnostic
Post-treatment
Remission
14M
7SS
14L
7SS

Supplementary Figure 1



Supplementary Figure

A



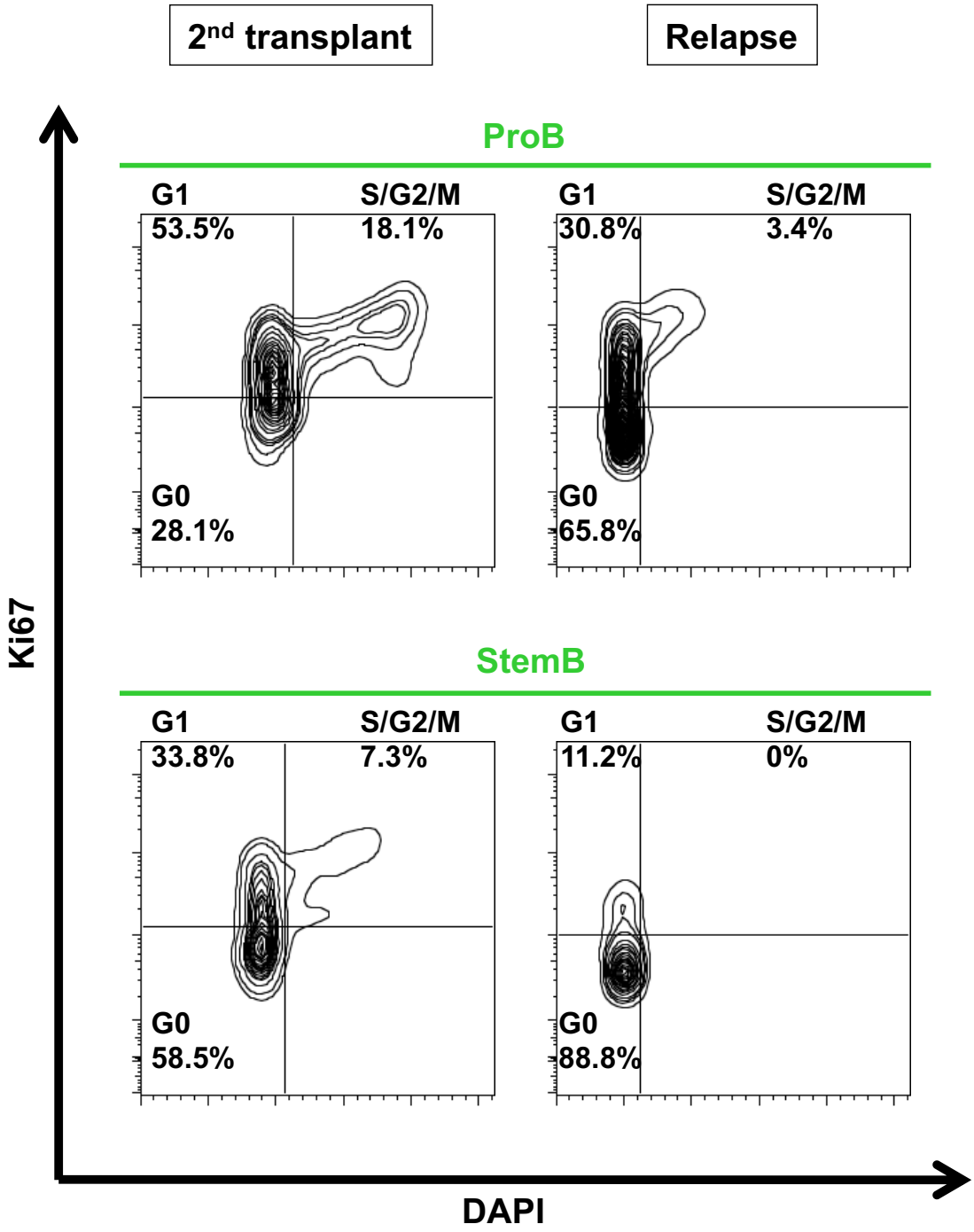
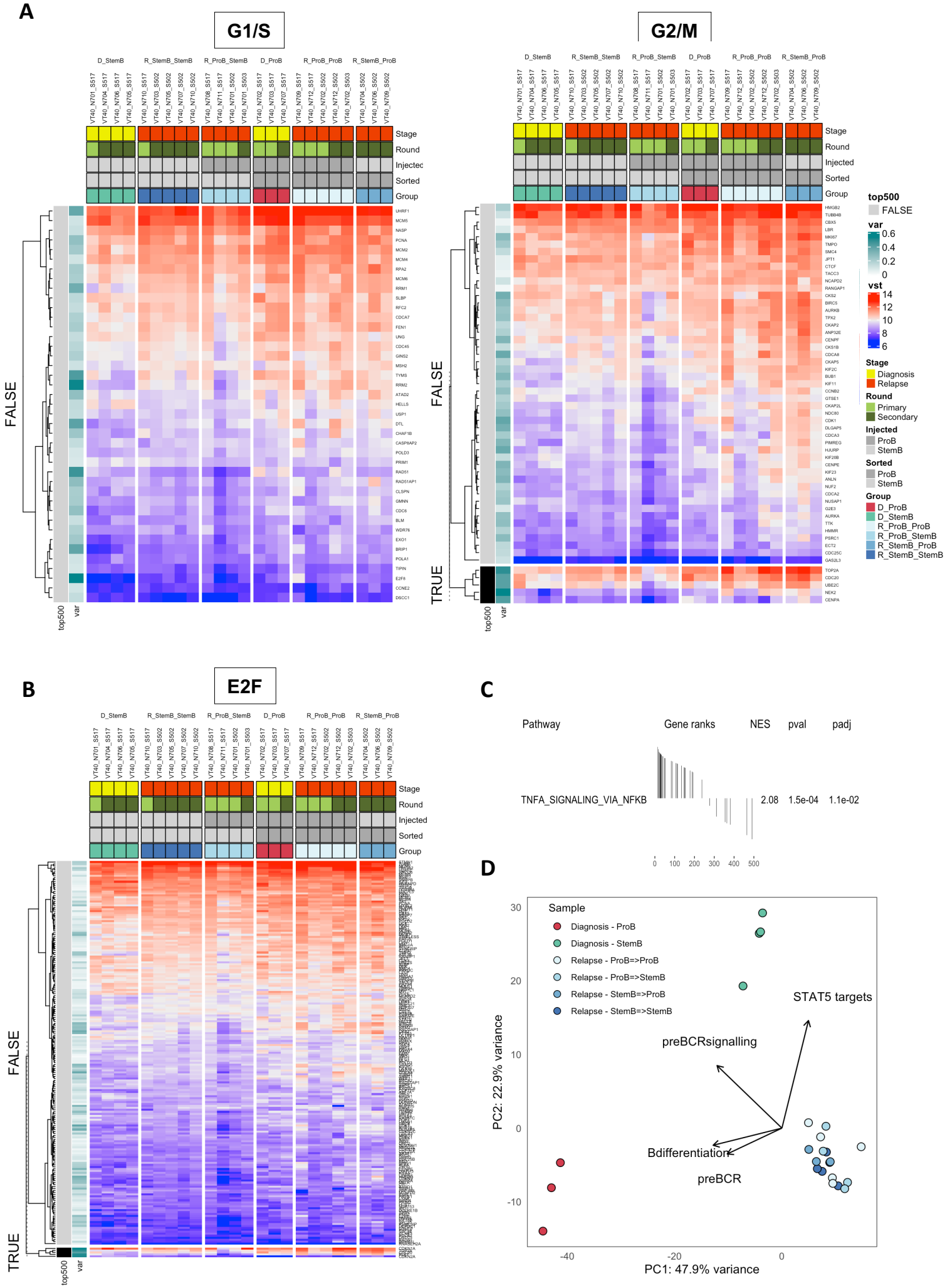
*** Ctrl: BCR-ABL status of non-Down's associated ALL**

	Ctrl *	Pt. 4
HSC	0	0
Stem/B	100	0
proB	100	100

100 cells/sample scored

Supplementary Figure 2

Cell cycle Analysis:



Supplementary Figure 3