

1

2

Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia

3

4

5

6

Christopher C. Oakes¹, Rainer Claus^{1,5}, Lei Gu^{1,2}, Yassen Assenov¹, Jennifer Hüllein⁶,

7

Manuela Zucknick³, Matthias Bieg², David Brocks¹, Olga Bogatyrova¹, Christopher R.

8

Schmidt¹, Laura Rassenti⁸, Thomas J. Kipps⁸, Daniel Mertens^{4,9}, Peter Lichter^{4,11},

9

Hartmut Döhner⁹, Stephan Stilgenbauer⁹, John C. Byrd¹⁰, Thorsten Zenz^{6,7} and

10

Christoph Plass^{1,11}

11

12

13

Affiliations: Division of ¹Epigenomics and Cancer Risk Factors, ²Theoretical

14

Bioinformatics, ³Biostatistics and ⁴Molecular Genetics, The German Cancer Research

15

Center (DKFZ), Heidelberg, Germany; ⁵Department of Medicine, University of Freiburg

16

Medical Center, Freiburg, Germany; ⁶Department of Translational Oncology, National

17

Center for Tumor Diseases (NCT) and German Cancer Research Center (DKFZ),

18

Heidelberg, Germany; ⁷Department of Medicine V, University of Heidelberg, Heidelberg,

19

Germany; ⁸Department of Medicine, University of California at San Diego Moores

20

Cancer Center, La Jolla, USA; ⁹Department of Internal Medicine III, University of Ulm,

21

Ulm, Germany; ¹⁰Division of Hematology, The Ohio State University, Columbus, USA;

22

¹¹The German Cancer Consortium.

23 **Corresponding Author:**

24 Christoph Plass

25 German Cancer Research Center (DKFZ),

26 Division of Epigenomics and Cancer Risk Factors,

27 Im Neuenheimer Feld 280, 69120 Heidelberg, Germany,

28 Tel. +49 6221-423330,

29 Fax +49 6221-423359,

30 E-mail: c.plass@dkfz.de

31

32 **Running Title:** Coevolution of epigenetics and genetics in CLL

33 **Key words:** DNA methylation; heterogeneity; coevolution; allele-specific methylation;

34 chronic lymphocytic leukemia

35

36 **Abbreviations list:**

37 CLL, chronic lymphocytic leukemia

38 CNA, copy number alteration

39 ASM, allele-specific methylation

40 BS-seq, bisulfite sequencing

41 SNP, single nucleotide polymorphism

42 WGBS, whole-genome bisulfite sequencing

43 MH, methylation heterogeneity

44 EPM, epipolymorphism

45 AML, acute myeloid leukemia

46 NBC, naïve B cell

47 ncsMBC, non-class-switched memory B cell

48 csMBC, class-switched memory B cell

49 LN, lymph node

50 PB, peripheral blood

51 SHM, somatic hypermutation

52

53 **Notes:**

54 This work was supported in part by The Helmholtz Association, from the DKFZ–
55 Heidelberg Center for Personalized Oncology (DKFZ-HIPO), the German Federal
56 Ministry of Education and Research CancerEpiSys network_(BMBF 031 6049C) and the
57 Virtual Helmholtz Institute (VH-VI-404). D.B. is has a stipend form the Geman Israeli
58 Helmholtz Graduate School, R.C. is funded by the German Cancer Aid through a Max
59 Eder Stipend, T.Z. is funded by the German Cancer Aid through a Stiftungsprofessur,
60 C.O. is a recipient of a post doctoral fellowship from the Leukemia and Lymphoma
61 Society.

62

63 **Conflict of interest disclosure statement:**

64 The authors declare no competing financial interests.

65 ABSTRACT:

66 Although clonal selection by genetic driver aberrations in cancer is well documented, the
67 ability of epigenetic alterations to promote tumor evolution is undefined. We used 450k
68 arrays and next-generation sequencing to evaluate intra-tumor heterogeneity and
69 evolution of DNA methylation and genetic aberrations in chronic lymphocytic leukemia
70 (CLL). CLL cases exhibit vast inter-patient differences in intra-tumor methylation
71 heterogeneity, with genetically clonal cases maintaining low methylation heterogeneity
72 and up to 10 percent of total CpGs in a monoallelically methylated state. Increasing
73 methylation heterogeneity correlates with advanced genetic subclonal complexity.
74 Selection of novel DNA methylation patterns is observed only in cases that undergo
75 genetic evolution, and independent genetic evolution is uncommon and is restricted to
76 low-risk alterations. These results reveal that although evolution of DNA methylation
77 occurs in high-risk, clinically-progressive cases, positive selection of novel methylation
78 patterns entail co-evolution of genetic alteration(s) in CLL.

79 SIGNIFICANCE:

80 Epigenetic alterations are pervasive in cancer and continually develop during disease
81 progression; however, the mechanisms that promote changes in the tumor epigenome at
82 large are currently undefined. The current work provides insight into the coevolution of
83 genetic and epigenetic aberrations and highlights the influential role of genetic
84 aberrations in the selection of novel methylation patterns.

85 INTRODUCTION:

86 The impact of genetic events on the development and progression of cancer has
87 been clearly demonstrated through the use of murine genetic tumor models and through
88 the association of recurrent mutations and genomic aberrations with clinical outcome.
89 Epigenetic differences are vast between tumor and perceived normal tissues, as well as
90 between patients, typically involving thousands of loci in a particular genome (1).
91 Epigenetic patterns between various normal cell types are highly divergent, and are key
92 in determining cell phenotypes and function (2, 3) . Although several onco- and tumor-
93 suppressor genes are found to have recurrently altered epigenetic states in tumors
94 which contribute to the cancer cell phenotype; a direct, causative role for the bulk of
95 epigenetic alterations is unclear. Recent tumor genome sequencing efforts have
96 uncovered mutations affecting numerous genes with known epigenetic functions in
97 cancer (reviewed by (4)), which further supports an important role for epigenetics in
98 cancer development.

99 Evolution and resulting genetic tumor heterogeneity is currently under
100 investigation for many malignancies since it may explain acquired resistance to
101 therapies. Pronounced intra-tumor genetic variation has been recently appreciated for
102 solid tumors (5-7), acute leukemias (8, 9) and chronic lymphocytic leukemia (CLL) (10,
103 11). In comparison to other cancers, CLL offers several advantages to study epigenetic
104 heterogeneity and evolution of tumor cell populations. Firstly, CLL is a malignancy that
105 possesses a mature, differentiated cellular phenotype that is epigenetically stable
106 throughout disease course, even following treatment (12). CLL tumor samples can be
107 obtained at near complete purity, and allow for the assignment of tumor sub-populations
108 to the original founder cell via the unique rearrangement of the B cell receptor. Lastly,
109 epigenetic patterns in CLL are consistent between peripheral blood and lymph node
110 compartments (12), allowing for the overall tumor cell population to be represented upon

111 sampling. Furthermore, evolution of genetic alterations in CLL is found to occur in
112 patients with poor prognostic markers and to be associated with inferior outcome (13).

113 Epigenetic alterations, such as DNA methylation, have the potential to add
114 complexity to the tumor cell population. Loss of epigenetic stability resulting in tumor
115 heterogeneity has been recently described to frequently occur in cancer (14, 15).
116 Studies of the CLL methylome have revealed an abundance of genes and other genomic
117 regions that display altered DNA methylation states (16, 17), including methylation
118 markers of high prognostic significance (18, 19). Despite the high frequency and
119 importance of epigenetic alterations, the contribution of DNA methylation patterns to
120 heterogeneity and evolution of tumor cell populations, and their relationship to genetic
121 evolution, is currently undefined.

122

123 RESULTS:

124 *CLL retains a large quantity of allele-specific methylation:*

125 Global DNA methylation was evaluated in 68 CLL samples and 11 healthy donor
126 B and T cell samples using Illumina human 450k BeadChip analysis. All samples were
127 purified to >99% by CD19+ or CD3+ selection for B or T cells, respectively. To mitigate
128 the influence of allele- and sample-specific variation in genomic sequence, all probes
129 overlapping non-unique sequences, SNPs and sample-specific copy number alterations
130 (CNAs) were removed from all 450k methylation profiles (see Methods). Though all CLL
131 and healthy donor samples display an enrichment of CpG methylation values in the
132 ranges of 0-20% (mainly CpG islands) and 80-100% (mainly gene body, intergenic
133 CpGs) as observed previously (17), CLL samples display a distinct 3rd peak of
134 intermediate methylation values centered around 50% (Fig.1a). The prominence of this
135 peak is highly variable between CLL samples and it is not observed in healthy donor B
136 or T cell samples. As diploidy is largely maintained in the genome of CLL cells (11), we

137 hypothesized that the intermediate peak may be the result of allele-specific methylation
138 (ASM). To test this, we performed bisulfite sequencing (BS-seq) targeting differentially
139 methylated regions of imprinted gene clusters as well as non-imprinted regions where
140 intermediately methylated (40-60%) CpGs were identified. Twenty-eight amplicons were
141 sequenced, including two imprinted domains as controls, in 20 CLL and 4 healthy donor
142 B cell samples with a median read depth of ~3800 reads. Average CpG methylation
143 determined by BS-seq was highly correlated with 450k β -values ($R^2=0.93$, Supplemental
144 figure 1). Twenty-three amplicons contained sufficient SNP frequency to assign alleles.
145 All imprinted CpGs demonstrated a difference of >75% methylation between alleles, thus
146 this value was used for the definition of ASM in other amplicons (Fig.1b). ASM can be
147 readily observed in CLL samples. To determine the overall ASM composition of the
148 intermediate peak on 450k profiles, 450k methylation values were plotted in comparison
149 to the methylation difference between alleles in the 10 CLL samples most prominently
150 displaying the intermediate peak (Fig.1b). This comparison reveals that 85% of 450k
151 values between 40-60% methylation (in non-imprinted regions) are monoallelically
152 methylated in these samples, demonstrating that the bulk of the CLL-specific
153 intermediate peak results from ASM. Although healthy donor lymphocyte samples show
154 values between 40-60%, only 0.4% of non-imprinted CpGs in healthy B cells exhibit
155 ASM. Analysis of the patterns of CLL-specific ASM reveals that neighboring CpGs
156 possess ASM on opposite alleles at random within individual amplicons (Supplemental
157 Figure 2). This is in contrast to imprinted regions where methylation always occurs solely
158 on the same allele (in both healthy and CLL cells), indicating that the majority of CLL-
159 specific ASM does not signify imprinting. This complex pattern of stable allelic
160 methylation has been suggested to likely occur via active demethylation (20). Moreover,
161 this feature also suggests that ASM in CLL may be distinct from the large partially
162 hypomethylated domains observed in other cancers (14, 21).

163 Genomic features associated with allele-specific CpG methylation (ASM-CpGs)
164 in CLL were analyzed by 450k arrays in the 10 CLL samples in which ASM was most
165 prominently observed. On average, only 20% of ASM-CpGs are found within CpG
166 islands and 31% in the vicinity of gene transcriptional start sites, and thus are more
167 similar in their genomic distribution to CpGs generally found to be fully methylated than
168 to unmethylated (Fig.1c). Indeed, 78% of CLL ASM-CpGs are fully methylated in to
169 healthy donor B cells, suggesting that ASM in CLL mostly results from allele-specific
170 methylation loss (Supplemental figure 3). This bias toward allele-specific loss of
171 methylation is consistent in comparisons to other B cell subtypes, including naïve CD5+
172 and memory-type B cells. Furthermore, ASM does not occur in patient-matched non-CLL
173 leukocytes (Supplemental figure 4). In contrast to CpGs in low or high methylation
174 ranges, the allele-specific methylation state of individual CpGs shows a very low (2.7%)
175 recurrence in CLL samples (Fig1.c). Although the bulk of ASM appears to occur by
176 chance, some ASM may recur non-randomly between samples (Supplemental figure 5).
177 A gene ontology survey of all genes enriched for ASM (>25% of CpGs/gene equaling an
178 average of ~10% of genes annotated per gene ontology group) revealed no significant
179 enrichment of ontology terms. Of the 2.7% recurrent CpGs, 28% are located within
180 known imprinted regions and 58% also display ASM in healthy B cells. After censoring
181 these CpGs, only 0.4% of overall ASM in CLL is recurrent and potentially disease-
182 specific.

183 The prevalence of ASM-CpGs in 450k profiles is highly variable between
184 individual CLL samples (Fig1.a,d). To estimate the levels of genome-wide ASM, the
185 proportion of enriched intermediate CpG methylation values was determined by
186 extrapolating a hypothetical curve connecting fully methylated and unmethylated
187 distributions (see Methods) (Fig.1d). Using this method, we estimate that genomic
188 monoallelic methylation ranges broadly from 2-10% of total CpGs in CLL (Fig.1e).

189 Healthy lymphocytes are estimated to possess <1% monoallelic methylation, consistent
190 with other genome-wide assessments (22, 23). To validate and further explore ASM on a
191 genome-wide level, ASM was assessed in whole-genome bisulfite sequencing (WGBS)
192 data of two CLL samples and three healthy B cell subtypes (17)). The prevalence of
193 ASM-CpGs was found to be approximately 6-8 fold higher in CLL samples relative to
194 healthy B cell subtypes (Supplemental figure 6). Furthermore, the number of ASM-CpGs
195 determined by WGBS is closely proportional to the estimated amount by 450k analysis
196 in the different CLL samples (Fig.1e). In comparing 450k ASM estimations in other
197 cancers (17, 24-27), CLL retains 3-5 fold more ASM (Fig1.f). Together these results
198 suggest that ASM in CLL is firstly due to a monoallelic loss of methylation prior to or
199 during the establishment of the CLL founder clone, followed by high fidelity maintenance
200 methylation which preserves methylation patterns *in cis* throughout subsequent
201 generations of cells.

202

203 *Methylation heterogeneity in CLL:*

204 To investigate the basis for the high degree of variation in ASM between CLL
205 samples, we hypothesized that the degree of ASM reflects intra-sample heterogeneity of
206 DNA methylation patterns. In a diploid cell, CpG methylation values are restricted to
207 three states (methylated, unmethylated and monoallelically methylated). If a population
208 of cells maintains a stable, clonal pattern of methylation, values derived from a sample
209 containing large numbers (usually $>1.0 \times 10^7$) of cells will also be restricted to these three
210 discrete ranges of CpG methylation values. As all CpGs found within CNAs have been
211 removed from the analysis, methylation values that occur between these discrete ranges
212 can only be caused by a disparate CpG methylation state between cells within the
213 sample. The total amount of CpGs that fall outside the expected ranges can thus be
214 used to estimate the overall level of methylation heterogeneity (MH) in a given sample.

215 This approach of elucidating intra-sample heterogeneity has been used previously in
216 conjunction with the HELP genome-wide methylation assay (28). Here, intra-sample MH
217 is calculated by summing all values between 20-80% methylation subtracted by the
218 amount of estimated genomic ASM (see Methods). Figure 2a displays the area of the
219 450k methylation value density plots used to define MH in two CLL samples showing
220 different levels of heterogeneity and in healthy donor B and T cell samples. MH values
221 for all samples are displayed in figure 2b. Due to the polyclonal nature of healthy B and
222 T cell populations, healthy donor lymphocyte samples would be anticipated to display
223 methylation heterogeneity, as subtypes of B and T cells exhibit distinct, genome-wide
224 patterns (17, 29). Indeed, healthy donor B and T cell samples display a relatively higher
225 level of methylation heterogeneity. B cells extracted from lymph nodes display higher
226 MH levels compared to peripheral blood B cells, likely due the high degree of B cell
227 diversification that occurs within germinal centers. Interestingly, MH values in CLL are
228 not normally distributed (Anderson-Darling test, $P < 0.001$), with a group of cases
229 clustering below the median (12.5%) level of MH. The non-normal distribution and
230 median value is comparable to an additional CLL 450k dataset (17) ($P < 0.001$). For this
231 reason, this median MH value is used to distinguish low and high MH groups for
232 subsequent analysis.

233 To confirm the accuracy of methylation heterogeneity estimations from 450k
234 profiles, we used BS-seq to determine the intra-sample heterogeneity of methylation
235 patterns in CLL and healthy B cell samples. For this, we employed the calculation of
236 epipolymorphism (EPM) (15), which is a measurement of the observed consistency of a
237 given pattern of methylation within a small defined region of neighboring CpGs (3-6
238 CpGs) versus the expected, random pattern. Low EPM values indicate that methylation
239 patterns are similar between cells in a population, whereas elevated EPM values reflect
240 higher heterogeneity. We calculated EPM from the BS-seq data generated on 20 CLL

241 and 4 healthy donor B cell samples. Healthy donor B cell samples demonstrate a low
242 degree of pattern consistency, with all possible methylation states represented in
243 proportions that would mostly be expected by chance (Fig.2c). In contrast, most CLL
244 samples demonstrate a higher degree of pattern consistency, and, in some amplicons,
245 only a single dominant pattern (epi-allele) per allele. These consistent methylation
246 patterns are observed despite highly discordant methylation existing between
247 neighboring CpGs and between alleles (see also supplemental figure 2b). Correlation of
248 intra-sample MH values with the average EPM across 25 amplicons reveals a general
249 agreement between the two methods ($R^2=0.86$), although MH evaluations by 450k
250 slightly underestimates the high intra-sample heterogeneity found by EPM in healthy
251 donor samples (Fig.2d).

252 Following confirmation of MH estimations from 450k data, we firstly asked
253 whether the level of genomic ASM is dependent on the amount of MH in a given CLL
254 sample. ASM and MH exhibit a strong inverse correlation in CLL ($R^2=0.66$,
255 Supplemental Figure 7). By definition, the existence of ASM requires an allelic CpG
256 methylation pattern to be highly consistent in a given population of cells (to fulfill the
257 criteria of a 75% methylation difference between alleles). Therefore, it is intuitive that
258 higher levels of overall MH reflects lower levels of ASM, and indicates that variable
259 methylation of ASM-CpG partially contributes to the overall level of MH in a sample.
260 Despite AML and CLL possessing similar levels of MH (Fig.2b), AML exhibits a much
261 lower level of ASM. This implies that ASM and MH are not merely two measures of the
262 same underlying phenomenon, and thus the high level of ASM in CLL is a distinctive
263 feature of the disease. Analysis of solid tumor data yields consistently higher overall MH
264 levels in comparison to CLL and AML. As the estimation of MH is highly influenced by
265 sample purity, it is likely that the true levels of heterogeneity between tumor cells are
266 overestimated in these samples. Absolute tumor cell content in solid cancers ranges

267 from 30-90% (30); however, glioblastomas possess >90% tumor nuclei in most samples
268 and display higher MH than all CLLs investigated (Supplemental figure 7). Together,
269 these results reveal that CLL exhibits a high level of genomic ASM relative to other
270 leukemias and solid tumors, and that this distinctive feature is facilitated by - but is not
271 specifically a result of - a low overall level of heterogeneity in the disease.

272 Next we investigated if MH is associated with disease-related factors, such as
273 prognostic indicators and patient outcome. Firstly, we compared various disease
274 markers of high prognostic significance, including *IGHV* mutation status (31), *ZAP70*
275 methylation (19) and cytogenetic profiling (32). Patients with an unmutated *IGHV* gene,
276 unmethylated *ZAP70* and/or high-risk cytogenetics, including deletion of 11q and 17p,
277 are generally associated with a more aggressive disease course. CLLs with above
278 median MH are more frequently *IGHV* unmutated and have low *ZAP70* methylation
279 (Table 1). Samples that were taken post-therapy also are found to possess high MH
280 more frequently than samples from non-treated patients. However, it is problematic to
281 attribute treatment as a direct cause of high MH, as high MH is associated with poor
282 prognosis and thus a greater likelihood of treatment. Indeed, untreated CLL patients
283 displaying an above median MH prior to therapy show a significantly reduced ($P=0.006$)
284 time from sampling to their first treatment (Fig.2e). This suggests that epigenetic
285 heterogeneity in the pre-treatment window is associated with a more aggressive disease
286 course.

287

288 *The relationship between epigenetic and genetic heterogeneity:*

289 Next we tested if the MH correlates with genetic heterogeneity in CLL samples.
290 To assess genetic heterogeneity, we postulated that biologically significant subclonal
291 populations would be identified by genomic events that have been shown to be relevant
292 to CLL biology. Thus, we assessed in each CLL sample: 1) the total number and

293 proportion of rearranged/mutated *IGHV* sequences by qPCR, Sanger and next-
294 generation sequencing approaches; 2) the frequency of recurrent somatic mutations in
295 the exons of *TP53*, *NOTCH1*, *SF3B1*, *MYD88*, *KRAS*, and *BRAF* by high-coverage 454-
296 based sequencing; 3) the proportional copy number of large (>1 Mb) genomic
297 aberrations by a non-biased, genome-wide approach derived from 450k arrays (24) and
298 by targeting recurrent CNAs in chromosomes 11, 13 and 17 using Taqman qPCR.
299 Finally, FISH and karyotype data were also used to establish whether common CNAs
300 were mono or biallelic. Using these quantitative data, the clone size that each mutation
301 and/or CNA represents was assigned in each sample. To designate a single value of
302 genetic heterogeneity to each sample, we identified from all available genetic data the
303 mutation and/or CNA clone size that would yield the most heterogeneous ratio of the two
304 largest clones. This value is termed here the genetic clone ratio. Using this approach, 66
305 of 68 CLL samples were assigned a genetic clone ratio (Supplemental table 1). Figure
306 3a illustrates the determination of the genetic clone ratio in two CLL samples. We
307 observe a strong relationship between MH and genetic heterogeneity, with higher MH
308 values observed with increasingly heterogeneous genetic clone ratios (Fig.3b;
309 $P < 0.0001$). Samples scored as biclonal (more than one primary founder CLL population
310 detected by *IGHV* rearrangements) were assessed separately and were found to have
311 high levels of MH. These data indicate that intra-sample MH is connected to the degree
312 of genetic diversification and relative proportions of subclonal populations.

313 To further investigate the relationship between epigenetic and genetic
314 heterogeneity, we focused on 28 CLL cases where samples were taken at two or more
315 time points (median difference of 29 months, range 12-113). The mutation and/or CNA
316 clone size for each aberration per sample was determined. The degree of change
317 between the timepoints for each case was defined by the mutation/CNA showing the
318 greatest difference. Representative CLL cases showing <20% (no/low change) or >50%

319 (large changes) in genetic clone ratios are displayed in figure 4a and b, respectively. The
320 difference in overall methylation was measured by calculating the correlation between
321 timepoints using the top 40k most variable probes between timepoints in all serial cases.
322 CLL cases without genetic evolution demonstrate consistent methylation between
323 timepoints, whereas cases that show high genetic evolution also show widespread
324 methylation changes over time.

325 In total, 13 of the 28 serial cases were observed to undergo a genetic change of
326 >20% (Fig.5a,b). By defining a difference between groups by both the number of CpGs
327 that differ by >10% and the R^2 -value (see Supplemental figure 8 for a detailed
328 description of group dichotomization), 9 of the 13 cases display evolution of methylation
329 as defined by more than 5×10^3 differentially methylated CpGs and $R^2 < 0.95$. The genetic
330 aberrations that are observed to evolve co-dependently with methylation involve a
331 subset of recurrent mutations and/or CNAs. The majority of these aberrations (i.e. those
332 involving *TP53*, *SF3B1*, *BRAF*, del11q23, del17p13, etc.) have been previously
333 described as subclonal cancer driver mutations that are frequently associated with
334 genetic evolution (11). Decrease or extinction of some mutations/CNAs is found to
335 occur, indicating that a potential hierarchy of aberrations exists between subclonal
336 populations. The four cases that do not show evolution of methylation, yet show a >20%
337 change in genetic clone ratio, specifically involve changes detected solely at the 13q14
338 locus, hinting that some aberrations may not be linked to methylation evolution.
339 Epigenetic and genetic changes are highly co-dependent (Fisher's exact test, $P < 0.001$),
340 as widespread epigenetic evolution independent of genetic evolution is not observed.

341

342 *Prediction and outcome of methylation evolution:*

343 We next investigated if evolution of DNA methylation is associated with
344 prognostic indicators or with specific genetic markers. Comparing the 9 serial cases that

345 showed methylation evolution versus the 19 cases that showed no/low evolution, we find
346 that a significant enrichment of *IGHV* unmutated and low *ZAP70* methylated cases
347 ($P=0.002$, Table 2). Intriguingly, those cases that showed a high level of MH in early
348 sample timepoints predicted the occurrence of evolution ($P=0.002$), supporting the
349 notion that high MH may result from active evolution. Methylation evolution is also
350 associated with intervening treatment, as 8/9 evolving (vs. 7/19 non-evolving) cases
351 received treatment between timepoints ($P=0.01$); however, based on the finding that
352 high MH predicts a shorter time to treatment (Fig.2e), it is likely that evolution provokes
353 treatment in at least an equal manner to treatment inducing evolution. The only mutation
354 or CNA that was significantly associated with predicting methylation evolution was *TP53*
355 ($P=0.03$), although the general low frequency of mutations in CLL necessitates a larger
356 cohort of evolving cases for further testing. Overall, the presence of a subclonal
357 mutation/CNA (<80% clone size) predicted methylation evolution ($P=0.04$); whereas the
358 presence of a clonal mutation did not, mirroring the findings of predicting genetic
359 evolution (11).

360 Next, we tested if the response to first-line therapy by comparing the presence of
361 methylation evolution with the duration of the event-free time window following first-line
362 therapy. Treatment and death were included as post-therapy events. All patients
363 included were previously untreated upon first sampling and subsequently treated with
364 purine analog and/or alkylating therapy (Supplemental table 1). Patients exhibiting
365 methylation evolution experienced post-therapy events in a substantially shorter time
366 than those lacking evolution (Fig.5c, median=9 vs. 110 mo., $P=0.0001$). Together, these
367 observations demonstrate an association between methylation evolution and poor
368 prognostic and genetic indicators, as well as to a lack of a durable response to therapy
369 | and a more aggressive disease course.

370

371 DISCUSSION:

372 CLL generally exhibits a remarkable stability of DNA methylation. Combining the
373 findings of others (12) with our findings, we demonstrated that CLL tumor populations
374 maintain a precise overall pattern of DNA methylation for many years of disease course.
375 Furthermore, as near-clonal patterns of methylation can be found in the cells of some
376 patients, a perfect maintenance of methylation states must occur from the initial,
377 founding epigenetic patterns associated with disease transformation. In these highly
378 stable clones, the vast amount of CpG methylation that occurs only on one allele in
379 nearly all cells is likely a simple reflection of the methylation status of the original founder
380 clone. CLL arises in a relatively mature cell type that has some hallmarks of memory-
381 type B cells, which may contribute to its stability phenotype compared to some other
382 leukemias. It is tempting to draw a parallel between the general indolent nature of the
383 disease and the extreme stability of the epigenome in some patients. Here, we also
384 show that in contrast to the high stability of DNA methylation generally observed in the
385 disease, a subset of cases demonstrate elevated levels of methylation heterogeneity.
386 Above median levels of methylation heterogeneity are associated with poor prognostic
387 indications, a shorter time to treatment, and greater subclonal genetic diversification.

388 The association of *IGHV* mutation status and other prognostic markers with our
389 findings advocates the integration of DNA methylation heterogeneity and evolution,
390 along with associated genetic aberrations, into the established high/low risk subtype
391 model of CLL (Fig.6). In this integrated view, ASM occurs in the founder malignant cell
392 as a result of monoallelic loss of methylation associated with B cell maturation (17)
393 and/or transforming events. Establishment is also usually associated with the acquisition
394 of recurrent CLL founder mutation(s), such as trisomy12, MYD88, and others (11).
395 Highly stable, clonal CLLs, which are much less likely to co-evolve epigenetic and
396 genetic changes, are typically the *IGHV* mutated/*ZAP70* methylated subtype. These

397 cases generally exhibit low MH and require less immediate treatment. Evolution of
398 methylation is not observed to occur in the absence of newly acquired and actively-
399 selecting genetic aberrations. In a minority of cases, a genetic change can be detected
400 without an appreciable change in methylation. In these cases, the observed change in
401 4/4 patients is solely a change at the 13q14 locus, a common aberration in the low-risk
402 CLL subtype. CLL cases with above median MH, including all of those that display
403 methylation evolution, are associated with *IGHV* unmutated/*ZAP70* unmethylated
404 markers. In this high-risk disease subtype, increasing MH is associated with an
405 increasingly complex subclonal genetic architecture. In all cases that show methylation
406 evolution, a change in genetic architecture is observed. Evolving genetic aberrations in
407 this subset of cases involve known cancer driver genes, including *TP53*, *SF3B1*, *BRAF*,
408 etc.

409 How does coevolution of epigenetics and genetics occur? There are two main
410 (non-mutually exclusive) hypotheses (Fig.6). A) *Simultaneous acquisition*: a novel
411 mutation of a cancer driver gene is acquired in a cell which fundamentally alters the
412 biology of the cell in a way that involves changes to the epigenome. B) *Step-wise*
413 *acquisition*: involves a mechanism where firstly there exists a low level of epigenetic
414 instability producing variation within the CLL population. When a cell from this population
415 then acquires a novel cancer driver mutation, the variant methylation pattern of the
416 particular cell hitchhikes on the subsequent subclonal expansion. This expansion then
417 permits the detection of the altered methylation pattern that would otherwise only be
418 detectible on a single-cell level previous to the expansion.

419 Why are epigenetic and genetic changes associated? A possibility one must
420 firstly consider is that they are mechanistically unrelated. In the step-wise acquisition
421 scenario, it is possible that epigenetic drift occurs independently of the stochastic
422 acquisition of driver mutations. Another possibility is that they are mechanistically linked.

423 Associated genetic and epigenomic states have been observed in several other cancers,
424 including mutations in *IDH1/2* in gliomas (33) and myeloid malignancies (34), *H3F3A* in
425 glioblastomas (24) and *BRAF* in colorectal cancer (35). In most of these well described
426 associations, mutations occur in genes with defined roles in epigenetic pathways
427 (reviewed by (4)). However, a direct causative connection to epigenetic regulation
428 remains elusive. It stands to reason that many recurrent, high-impact mutations, not
429 known to directly involve epigenetic regulation, also involve epigenetic deregulation as a
430 part of their aberrant function. For example, the deletion of chromosome 17p is
431 associated with a loss of methylation at repetitive sequences in CLL (36). The most
432 judicious scenario places genetic events as the driving force behind the subsequent
433 evolution of a novel epigenetic state. However, one cannot exclude that primary changes
434 to the epigenome permit the acquisition of specific mutations, i.e. epigenetic silencing of
435 key tumor-suppressors which would otherwise have resulted in apoptotic cell
436 death/senescence (37). Epigenetic drift may endow a subset of cells within the
437 population with the eventual attributes needed to escape negative feed-back regulation
438 by tumor-suppressors, allowing for a driver mutation to occur. Here we observe that
439 epigenetic/genetic coevolution involves a spectrum of aberrations, implying a potentially
440 very broad and intricate inter-relationship between the genome and epigenome. Using
441 higher resolution techniques, future work will involve unraveling the relative contributions
442 of epigenetic versus genetic evolution to disease, and investigate whether monitoring
443 DNA methylation heterogeneity during disease course will benefit patients.

444

445 MATERIALS AND METHODS:

446 **CLL and healthy donor lymphocyte samples**

447 Clinical and biological characteristics of the 107 samples of CLL patients and healthy
448 donor controls used for DNA methylation analysis are shown in Supplemental Table 1.

449 CLL cases were selected to provide a balanced cohort for *IGHV* mutation status (28/68
450 <98% identity), treatment status (19/68 untreated, 19/68 treated post sampling, 30/68
451 treated before sampling), and treatment response to first line therapy (28
452 complete/partial response, 13 stable/progressive disease). Furthermore, samples were
453 enriched for the presence of informative somatic aberrations. Thus the cohort is not a
454 true representation of the general CLL population at large. FISH, *IGHV* and *ZAP70*
455 methylation analysis was done as previously described (19, 32, 38). All patients gave
456 informed consent.

457

458 **Isolation and purification of CLL and healthy lymphocytes**

459 All samples were obtained from whole blood then subjected to Ficoll-Isopaque density
460 centrifugation, and CD19+ B and CD3+ T cells were isolated by positive magnetic cell
461 separation (Miltenyi Biotec). Sorted cells were checked for purity by FACS with
462 CD19/CD20 for healthy control samples and CD19/CD20/CD5 for CLL samples (BD
463 Biosciences). Following sorting, all samples with a CD19/CD20/CD5 purity <98% were
464 subjected to additional sorting, and the average final purity of all sorted samples was
465 >99%. CLL samples with $>100 \times 10^6$ WBC/ul were not subject to purification. DNA was
466 extracted from purified cells using the Qiagen DNeasy kit (Qiagen) and quantified using
467 a ND-100 spectrophotometer (Thermo Scientific).

468

469 **DNA methylation analysis using 450k BeadChip arrays**

470 500 ng of high quality genomic DNA was bisulfite converted using the EZ DNA
471 Methylation gold Kit (Zymo Research). The Infinium methylation assay was carried out
472 as described previously (39). Data from the 450k Human Methylation Array were
473 normalized by the BMIQ method (40) using the RnBeads analysis software package
474 (41). Data are available at the European Genome-Phenome Archive

475 (EGAS00001000534). GenomeStudio (Illumina, Inc.) was used for CpG island and gene
476 segment annotation; repetitive sequence, segmental duplication, SNP and imprinted
477 DMR annotation was obtained from the UCSC genome browser, version hg19.
478 Estimation of genomic ASM from 450k methylation frequency plots was calculated by
479 firstly generating a hypothetical 3rd degree polynomial curve that estimates the
480 distribution without intermediate methylation values (i.e. from unmethylated and fully
481 methylated distributions) with smooth connections at fixed departure points (matching
482 the original function at these departure points in the first derivative). Estimated ASM is
483 the quantity of methylation values above the hypothetical curve and below the actual
484 density curve relative to all values analyzed (multiplied by 100 for scaling purposes).
485 Methylation heterogeneity (MH) was calculated by measuring the quantity of methylation
486 values below the hypothetical curve and between 20-80% methylation (again multiplied
487 by 100). This methylation window represents the range where the greatest difference
488 occurs between clonal CLL and healthy lymphocyte (polyclonal) samples. Different
489 variable and fixed MH window settings were tested and did not significantly change the
490 relative order of MH sample values or the association of MH versus outcome
491 (Supplemental figure 9). The reproducibility of estimated genomic ASM and MH values
492 was confirmed by testing two independent samples in two CLL cases, each sample was
493 independently isolated and purified (Supplemental figure 10). Because each CLL sample
494 may have a unique CNA profile, in addition to censoring all probes on chromosome arms
495 11q, 13q, 17p and 12p+q in all 450k profiles, any CNA >1Mb in size was censored in
496 sample-specific manner. Censoring was matched between serial samples. CNAs were
497 detected using an algorithm for quantitative CNA detection based on 450k probe
498 intensities (24). CLL 450k profiles were also censored for all non-unique sequences,
499 probes possibly containing SNPs, and sex chromosomes (totaling ~185K CpGs).
500 Additional CLL and healthy B cell 450k/WGBS data presented in figures 1, 2 and

501 supplemental figures 3, 6 & 7 were obtained from previously published work (17), acute
502 myleoid leukemia (25), glioblastoma (24, 26), renal clear cell carcinoma (26), colon
503 adenocarcinoma (27), lung adenocarcinoma (26).

504

505 **Analysis of ASM using WGBS**

506 WGBS data were obtained from ICGC (<http://icgc.org>). ASM-CpG were determined by
507 identifying heterozygous SNPs using the Bis-SNP algorithm (42) followed by determining
508 the allelic methylation ratio of each CpG within overlapping reads (minimum 8 reads per
509 allele). The calculation of the ASM to investigated CpG ratio required the ASM-CpG to
510 have a *P*-value smaller than $FDR \leq 0.05$ and a methylation difference of at least 75%.
511 CpGs were only considered if not overlapping a SNP. All CpGs were filtered which are
512 located in problematic regions (HISEQDEPTH, REPEAT_MASKER,
513 DUKE_EXCLUDED, and DAC_BLACKLIST; tracks obtained from the UCSC Genome
514 Browser, hg19).

515

516 **Targeted bisulfite sequencing and analysis**

517 Bisulfite-converted genomic DNA was amplified by standard PCR using barcoded
518 primers for patient sample identification. Primer sequences, SNPs and 450k probes
519 covered are listed in Supplemental table 2. Multiple PCR products from 12x2 samples
520 were pooled and sequenced using paired-end, 150bp reads on a MiSeq sequencer
521 (Illumina, Inc.). Median read depth per amplicon per patient was ~3800 high quality
522 reads. Debarcoded reads were analyzed simultaneously for methylation and genotype
523 using the Bis-SNP algorithm (42). Epipolymorphism (EPM) analysis was performed as
524 previously described (15) with modifications. To normalize EPM values derived from
525 amplicons with different numbers of CpGs and variable average methylation content,
526 firstly expected EPM values were generated by random simulation of methylation

527 patterns for amplicons containing 3-6 CpGs for average methylation ranges of 20-80%.
528 Spline curves derived from simulations were used to adjust EPM for average methylation
529 content of each amplicon in each sample using: $EPM = EPM_{\text{observed}} + (1 - EPM_{\text{expected}})$.
530 Amplicons with an average methylation <20% or >80% were excluded from EPM
531 analysis due to low complexity potential.

532

533 **The identification of genomic aberrations and the determination of genetic**
534 **heterogeneity in CLL samples**

535 Somatic genetic aberrations were assessed in 106 CLL samples. For each sample, the
536 sequence identity of the unique rearranged *IGHV* region was determined by genescan
537 qPCR followed by Sanger dye-terminator sequencing (38). Biclinality was defined by a
538 CLL sample exhibiting a minimum of 3 unique and fully recombined *IGHV* alleles, with a
539 minimum of two productive rearrangements. For samples with polyclonal chromatogram
540 profiles, PCR products were sequenced using MiSeq to determine the sequence and
541 proportion of subclones. The frequency of recurrent somatic SNVs in the exons of *TP53*,
542 *NOTCH1*, *SF3B1*, *MYD88*, *KRAS* and *BRAF* was determined by 454-sequencing (43)
543 (Roche). At least one mutation could be detected in 66/96 samples. All mutations were
544 considered to be heterozygous. The proportional copy number of large (>1 Mb) CNAs
545 was determined by a custom quantitative algorithm derived from 450k array raw data
546 (24). The proportional copy number of recurrent minimally-deleted regions (MDRs) in
547 chromosomes 11, 13 and 17 was further supported using Taqman qPCR. Eight primer-
548 probes were used to amplify various regions within each MDR and compared to eight
549 primer-probes positioned at various genomic positions not affected by CNAs in all
550 samples. CNAs could be detected in 88/96 samples. FISH data on chromosomes
551 6,8,11,12,13,14 and 17 were used to establish whether common CNVs were mono or
552 biallelic. In all, quantitative SNV/CNA data could be determined for 93/96 samples.

553

554

555 **Statistics**

556 Associations between MH, genetic heterogeneity and clinical features were assessed by
557 the Wilcoxon rank-sum test, Fisher exact test, or the Kruskal-Wallis test, as appropriate.

558 Correlation calculations were performed by Pearson product-moment correlation
559 coefficient (R^2). To test the significance of recurrence of ASM between samples we
560 constructed a test statistic which is the number of ASM-CpGs occurring in at least 8/10
561 samples, then an empirical P -value was calculated based on 10,000 permutations. Time
562 to event data were estimated by Kaplan-Meier analyses, and differences between
563 groups were assessed using the Mantel-Cox log rank test.

564

565

566 ACKNOWLEDGEMENTS:

567 We would like to thank for the excellent technical support and expertise at the German
568 Cancer Research Center (DKFZ) Genomics and Proteomics Core Facility. We are
569 grateful to Marion Bähr, Oliver Mücke, Monika Helf and Tatjana Stolz for technical
570 support and to Volker Hovestadt for helpful discussions. We thank David Lucas, Martina
571 Seiffert and Andrea Schnaiter for efficient distribution of samples and data.

572

573

574

575 REFERENCES:

- 576 1. Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and
577 translational implications. *Nat Rev Cancer*. 2011;11:726-34.
- 578 2. Broske AM, Vockentanz L, Kharazi S, Huska MR, Mancini E, Scheller M, et al. DNA
579 methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nat*
580 *Genet*. 2009;41:1207-U69.
- 581 3. Trowbridge JJ, Snow JW, Kim J, Orkin SH. DNA Methyltransferase 1 Is Essential for and
582 Uniquely Regulates Hematopoietic Stem and Progenitor Cells. *Cell Stem Cell*. 2009;5:442-9.
- 583 4. Timp W, Feinberg AP. Cancer as a dysregulated epigenome allowing cellular growth
584 advantage at the expense of the host. *Nat Rev Cancer*. 2013.
- 585 5. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor
586 heterogeneity and branched evolution revealed by multiregion sequencing. *The New England*
587 *journal of medicine*. 2012;366:883-92.
- 588 6. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life
589 history of 21 breast cancers. *Cell*. 2012;149:994-1007.
- 590 7. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution
591 inferred by single-cell sequencing. *Nature*. 2011;472:90-4.
- 592 8. Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA, et al. Genomic analysis of
593 the clonal origins of relapsed acute lymphoblastic leukemia. *Science*. 2008;322:1377-80.
- 594 9. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in
595 relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*.
596 2012;481:506-10.

- 597 10. Schuh A, Becq J, Humphray S, Alexa A, Burns A, Clifford R, et al. Monitoring chronic
598 lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal
599 evolution patterns. *Blood*. 2012;120:4191-6.
- 600 11. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, et al.
601 Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia. *Cell*.
602 2013;152:714-26.
- 603 12. Cahill N, Bergh AC, Kanduri M, Goransson-Kultima H, Mansouri L, Isaksson A, et al. 450K-
604 array analysis of chronic lymphocytic leukemia cells reveals global DNA methylation to be
605 relatively stable over time and similar in resting and proliferative compartments. *Leukemia*.
606 2013;27:150-8.
- 607 13. Stilgenbauer S, Sander S, Bullinger L, Benner A, Leupolt E, Winkler D, et al. Clonal
608 evolution in chronic lymphocytic leukemia: acquisition of high-risk genomic aberrations
609 associated with unmutated VH, resistance to therapy, and short survival. *Haematol-Hematol J*.
610 2007;92:1242-5.
- 611 14. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al.
612 Increased methylation variation in epigenetic domains across cancer types. *Nat Genet*.
613 2011;43:768-U77.
- 614 15. Landan G, Cohen NM, Mukamel Z, Bar A, Molchadsky A, Brosh R, et al. Epigenetic
615 polymorphism and the stochastic formation of differentially methylated regions in normal and
616 cancerous tissues. *Nat Genet*. 2012;44:1207-14.
- 617 16. Rush LJ, Raval A, Funchain P, Johnson AJ, Smith L, Lucas DM, et al. Epigenetic profiling in
618 chronic lymphocytic leukemia reveals novel methylation targets. *Cancer Res*. 2004;64:2424-33.

- 619 17. Kulis M, Heath S, Bibikova M, Queiros AC, Navarro A, Clot G, et al. Epigenomic analysis
620 detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat*
621 *Genet.* 2012;44:1236-42.
- 622 18. Kanduri M, Cahill N, Goransson H, Enstrom C, Isaksson A, Rosenquist R. Differential
623 Genome-Wide Methylation Profiles in Prognostic Subsets of Chronic Lymphocytic Leukemia.
624 *Haematol-Hematol J.* 2009;94:200-.
- 625 19. Claus R, Lucas DM, Stilgenbauer S, Ruppert AS, Yu LB, Zucknick M, et al. Quantitative
626 DNA Methylation Analysis Identifies a Single CpG Dinucleotide Important for ZAP-70 Expression
627 and Predictive of Prognosis in Chronic Lymphocytic Leukemia. *J Clin Oncol.* 2012;30:2483-91.
- 628 20. Ehrlich M, Lacey M. DNA hypomethylation and hemimethylation in cancer. *Advances in*
629 *experimental medicine and biology.* 2013;754:31-56.
- 630 21. Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, Pelizzola M, et al. Global DNA
631 hypomethylation coupled to repressive chromatin domain formation and gene silencing in
632 breast cancer. *Genome research.* 2012;22:246-58.
- 633 22. Fang F, Hodges E, Molaro A, Dean M, Hannon GJ, Smith AD. Genomic landscape of
634 human allele-specific DNA methylation. *P Natl Acad Sci USA.* 2012;109:7332-7.
- 635 23. Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J, et al. Base-Resolution Analyses of
636 Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. *Cell.*
637 2012;148:816-31.
- 638 24. Sturm D, Witt H, Hovestadt V, Khuong-Quang DA, Jones DTW, Konermann C, et al.
639 Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of
640 Glioblastoma. *Cancer Cell.* 2012;22:425-37.
- 641 25. Network CGAR. Genomic and epigenomic landscapes of adult de novo acute myeloid
642 leukemia. *The New England journal of medicine.* 2013;368:2059-74.

- 643 26. TCGA. The Cancer Genome Atlas Data Portal. [https://tcga-](https://tcga-datancinihgov/tcga/tcgaHome2jsp)
644 [datancinihgov/tcga/tcgaHome2jsp](https://tcga-datancinihgov/tcga/tcgaHome2jsp). 2013.
- 645 27. Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA, Fowler G, et al.
646 Comprehensive molecular characterization of human colon and rectal cancer. *Nature*.
647 2012;487:330-7.
- 648 28. De S, Shaknovich R, Riester M, Elemento O, Geng HM, Kormaksson M, et al. Aberration
649 in DNA Methylation in B-Cell Lymphomas Has a Complex Origin and Increases with Disease
650 Severity. *Plos Genet*. 2013;9.
- 651 29. Ji H, Ehrlich LIR, Seita J, Murakami P, Doi A, Lindau P, et al. Comprehensive methylome
652 map of lineage commitment from haematopoietic progenitors. *Nature*. 2010;467:338-U120.
- 653 30. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute
654 quantification of somatic DNA alterations in human cancer. *Nat Biotechnol*. 2012;30:413-21.
- 655 31. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V-H genes are
656 associated with a more aggressive form of chronic lymphocytic leukemia. *Blood*. 1999;94:1848-
657 54.
- 658 32. Dohner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L, et al. Genomic
659 aberrations and survival in chronic lymphocytic leukemia. *New Engl J Med*. 2000;343:1910-6.
- 660 33. Yan H, Parsons DW, Jin GL, McLendon R, Rasheed BA, Yuan WS, et al. IDH1 and IDH2
661 Mutations in Gliomas. *New Engl J Med*. 2009;360:765-73.
- 662 34. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, et al. Recurring
663 Mutations Found by Sequencing an Acute Myeloid Leukemia Genome. *New Engl J Med*.
664 2009;361:1058-66.

- 665 35. Weisenberger DJ, D Siegmund K, Campan M, Young J, Long TI, Faasse MA, et al. CpG
666 island methylator phenotype underlies sporadic microsatellite instability and is tightly
667 associated with BRAF mutation in colorectal cancer. *Nat Genet.* 2006;38:787-93.
- 668 36. Fabris S, Bollati V, Agnelli L, Morabito F, Motta V, Cutrona G, et al. Biological and clinical
669 relevance of quantitative global methylation of repetitive DNA sequences in chronic lymphocytic
670 leukemia. *Epigenetics.* 2011;6:188-94.
- 671 37. Hinoue T, Weisenberger DJ, Pan F, Campan M, Kim M, Young J, et al. Analysis of the
672 Association between CIMP and BRAF(V600E) in Colorectal Cancer by DNA Methylation Profiling.
673 *Plos One.* 2009;4.
- 674 38. Krober A, Buhler A, Kienle D, Benner A, Lichter P, Dohner H, et al. Analysis of VDJ
675 rearrangement structure and VH mutation status in chronic lymphocytic leukemia. *Blood.*
676 2001;98:358a-a.
- 677 39. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou LX, Shen R, et al. Genome-wide DNA
678 methylation profiling using Infinium (R) assay. *Epigenomics-Uk.* 2009;1:177-200.
- 679 40. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A
680 beta-mixture quantile normalization method for correcting probe design bias in Illumina
681 Infinium 450 k DNA methylation data. *Bioinformatics.* 2013;29:189-96.
- 682 41. Assenov Y, Muller F, Lutsik P, Walter J, Lengauer T, Bock C. Comprehensive Analysis of
683 DNA Methylation Data with RnBeads. <http://rnbeadsmpi-infmpgde>. 2013.
- 684 42. Liu YP, Siegmund KD, Laird PW, Berman BP. Bis-SNP: Combined DNA methylation and
685 SNP calling for Bisulfite-seq data. *Genome Biol.* 2012;13.
- 686 43. Hullein J, Jethwa A, Stolz T, Blume C, Sellner L, Sill M, et al. Next-generation sequencing
687 of cancer consensus genes in lymphoma. *Leuk Lymphoma.* 2013.

688 44. Edlmann J, Holzmann K, Miller F, Winkler D, Buhler A, Zenz T, et al. High-resolution
689 genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations.
690 Blood. 2012;120:4783-94.

691 TABLES:

692

693 Table 1: Comparison of methylation heterogeneity with patient characteristics
 694 & prognostic indicators:

	Low DNA methylation heterogeneity no. (%)	High DNA methylation heterogeneity no. (%)	<i>P</i>-value
<u>Patient characteristics:</u>			
Age at diagnosis (yr. \pm s.d.)	55.8 \pm 11.8	59.9 \pm 10.6	<i>n.s.</i>
Sex (Female)	10 (33)	15 (48)	<i>n.s.</i>
Pre-treatment (Yes)	6 (19)	13 (42)	0.014
<u>Prognostic indicators:</u>			
<i>IGHV</i> unmutated	8 (35)	11 (79)	<0.01
ZAP70 low methylation	7 (30)	9 (69)	0.024
Cytogenetics (NK, sole -13q)	14 (61)	9 (69)	<i>n.s.</i>

695

696 Table 2: Summary table of prognostic & genetic markers in serial CLL cases (timepoint 1):

	No/low evolution of DNA methylation <i>n</i> =19 (%)	High evolution of DNA methylation <i>n</i> =9 (%)	<i>P</i> -value
<u>Prognostic indicators:</u>			
<i>IGHV</i> unmutated	7 (37)	9 (100)	0.002
ZAP70 low methylation	7 (37)	9 (100)	0.002
High methylation heterogeneity	5 (26)	8 (89)	0.002
<u>Cytogenetics:</u>			
normal karyotype	4 (21)	2 (22)	<i>n.s.</i>
del 13q14	12 (63)	6 (67)	<i>n.s.</i>
del 11q23	4 (21)	1 (11)	<i>n.s.</i>
del 17p13	3 (16)	1 (11)	<i>n.s.</i>
del 6q	1 (5)	1 (11)	<i>n.s.</i>
trisomy 12	3 (16)	1 (11)	<i>n.s.</i>
<u>Genetic mutations:</u>			
<i>TP53</i>	3 (16)	5 (56)	0.03
<i>SF3B1</i>	4 (21)	3 (33)	<i>n.s.</i>
<i>NOTCH</i>	3 (16)	0	<i>n.s.</i>
<i>MYD88</i>	3 (16)	0	<i>n.s.</i>
<i>BRAF</i>	1 (5)	2 (22)	<i>n.s.</i>
<i>KRAS</i>	0	1 (11)	<i>n.s.</i>
<u>Mutation clone size:</u>			
clonal (>80%)	7 (37)	1 (11)	<i>n.s.</i>
subclonal (<80%)	5 (26)	6 (67)	0.04
any	11 (58)	7 (78)	<i>n.s.</i>

697

698 FIGURE LEGENDS:

699

700 Figure 1: Pronounced allele-specific methylation in CLL samples. (a) Frequency distribution of
701 CpG methylation values from 450k profiles in CLL and healthy donor B and T cell samples. CLL
702 display a prominent enrichment of methylation values centered around 50%. (b) A scatterplot
703 comparing 450k methylation versus the percent methylation difference between alleles
704 determined by BS-seq. CpGs from non-imprinted loci in CLL samples (blue dots), healthy B cells
705 (red dots), and imprinted loci (black Xs) are shown. The range of methylation difference defined
706 as allele-specific (>75% difference) is shown. (c) Genomic characteristics of CpGs separated into
707 unmethylated (0-20%), monoallelic (40-60%) and biallelically (80-100%) methylated CpGs in
708 clonal CLL samples. The proportion of CpGs associated with CpG islands, gene segments and
709 recurrence within each methylation range are displayed. (d) Methylation density plots of two
710 CLL samples representative of high (CLL44) and low (CLL112) monoallelic methylation, as well as
711 a healthy donor B and T cell sample, with the area used to estimate the overall proportion of
712 genomic ASM highlighted. (e) A comparison of ASM in whole-genome BS-seq (WGBS) with the
713 estimation by 450k. 450k methylation density plots of one healthy B cell sample and two CLL
714 samples analyzed by 450k are shown (top panels) along with the correlation between methods
715 (NBC, naïve B cell; ncsMBC, non-class-switched memory B cell). (f) Estimated ASM in all 68 CLL
716 and healthy donor lymphocyte samples. Individual values for CLL samples illustrated in figure 1d
717 are indicated. Analysis of downloaded 450k datasets (in grey) is included for comparison. CLL
718 (17); AML, acute myleoid leukemia (25); GBM, glioblastoma multiforme (24); Renal, renal clear
719 cell carcinoma (26); Colon, colon adenocarcinoma (27); Lung, lung adenocarcinoma (26).

720

721 **Figure 2: Variable intra-tumor heterogeneity of DNA methylation heterogeneity (MH) in CLL**
722 **samples.** (a) The proportion of 450k methylation values used to estimate the overall level of
723 DNA methylation heterogeneity in representative CLL and healthy donor samples. (b) MH values
724 show pronounced variation among CLL cases and collectively display lower MH than healthy
725 donor samples as well as other solid tumors. (LN, lymph node; PB, peripheral blood) (c) A
726 representative example of targeted allele-specific bisulfite-sequencing (surrounding the SNP
727 rs365605) showing mostly clonal (CLL21,44&86) and increasingly heterogeneous (CLL32,112)
728 methylation patterns among CLL samples. Despite disordered methylation states between
729 neighboring CpGs (horizontal), many CLL samples display mostly clonal patterns indicated by a
730 high proportion of identical epi-alleles (vertical). Epipolymorphism (EPM) and overall 450K MH
731 values are displayed, asterisks indicate ASM-CpGs. (d) Correlation between MH and the average
732 EPM for 25 targeted regions in 20 CLL and 4 healthy donor B cell samples demonstrates an
733 agreement between the two methods. (e) The duration of treatment-free survival from the time
734 of sampling to first therapy. CLL samples were segregated into two groups by the median MH
735 value of all samples.

736 **Figure 3: Epigenetic heterogeneity is associated with genetic heterogeneity in CLL samples.** (a)
737 Quantitative assessment of the mutation/CNA clone sizes for various aberrations in two CLL
738 samples representative of different levels of genetic heterogeneity. Clone sizes for various
739 detected somatic aberrations (grey) are displayed for CLL48 and CLL109. For CLL48, all variations
740 fall within a range consistent with a clonal sample population possessing monoallelic aberrations
741 at 13q14, 17p13 and *MYD88* and a single-copy gain of chromosome 12. For CLL109, individual
742 somatic variations occur at a frequency indicative of intraclonal diversification, with the
743 mutation clone size of the R625C mutation in *SF3B1* representing approximately an even ratio of
744 genetic clones. The most possible heterogeneous ratio of all mutations/CNAs is designated as

745 the genetic clone ratio for a given sample. (b) MH levels of all CLL samples versus the genetic
746 clone ratio. Biclonal samples are also displayed.

747

748 **Figure 4: Coordinated epigenetic and genetic evolution in 28 serial CLL cases.** (a)
749 Representative CLL cases showing no/low change ($\Delta < 20\%$) and (b) large change ($\Delta > 50\%$) in
750 genetic clone size are displayed. The time elapsed between sampling is displayed above each
751 panel. The mutation/CNA clone size determined for each aberration is shown for both
752 timepoints (above panels); error bars indicate standard deviation of technical replicates.
753 Recurrent CLL aberrations, defined by Edelmann *et al.* (44), are labeled in black, non-recurrent
754 CNAs in grey. Differences in clone size between timepoints that would represent a change of
755 $< 20\%$, $20\text{-}50\%$ and $> 50\%$ are illustrated by light blue, purple and pink areas, respectively. For
756 each sample, the methylation values of the overall 40k most variable CpGs are used to calculate
757 the Pearson correlation coefficient (R^2) between timepoints.

758

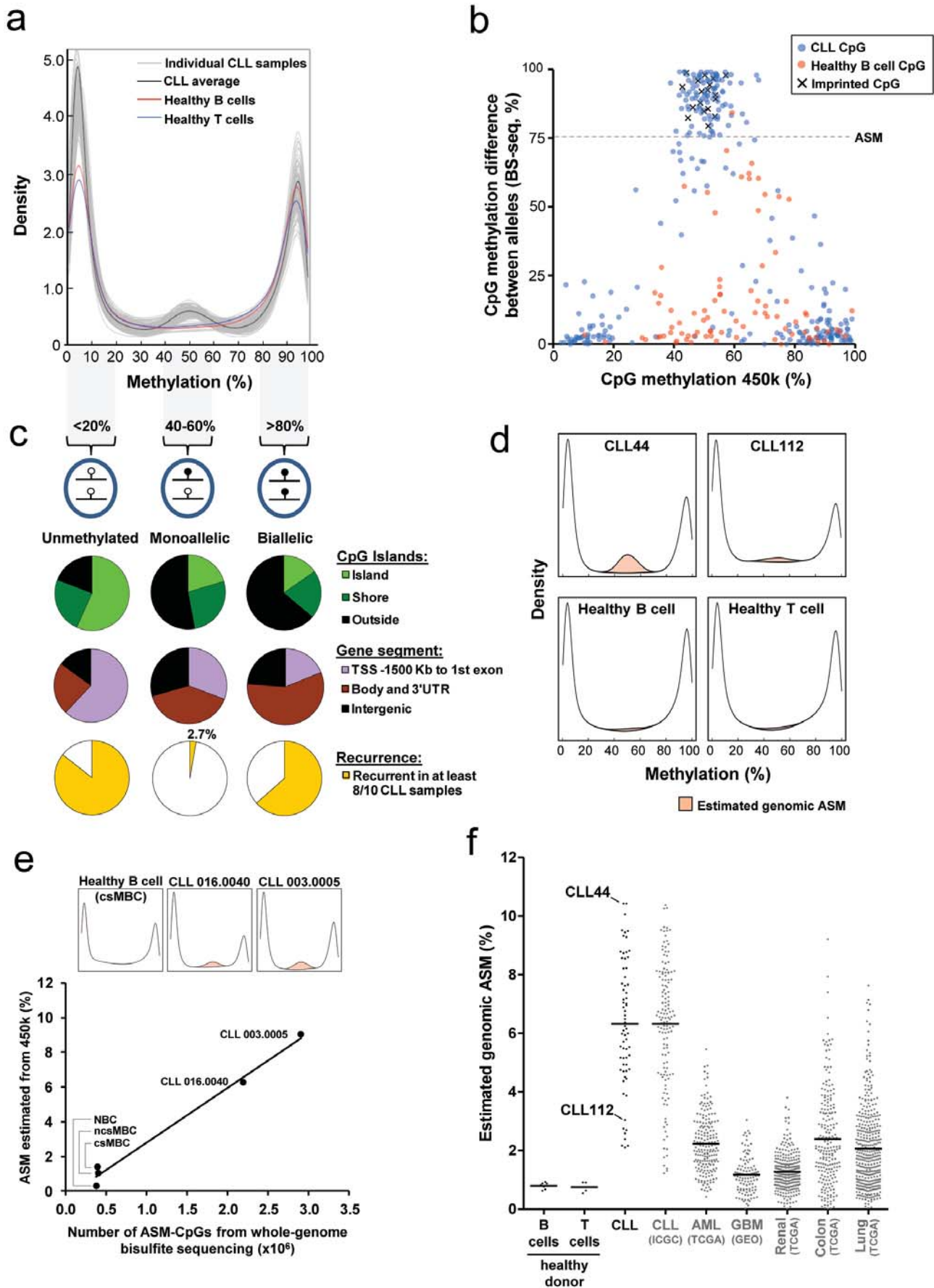
759 **Figure 5: Evolution of DNA methylation versus genetic aberrations and event-free survival**
760 **following first therapy.** (a) Correlation of genetic evolution (measured by the change in the
761 genetic clone ratio) with methylation evolution (measured by the number of differentially
762 methylated CpGs $\Delta > 10\%$) in 28 serial CLL cases. Cases that show no/low methylation or genetic
763 evolution (black dots) and co-evolving cases (red dots) are shown. Cases that show only genetic
764 evolution are colored grey. (b) A summary of methylation and genetic evolution in serial cases.
765 The change in methylation (measured by the number of differentially methylated CpGs and the
766 Pearson correlation) and the change in the genetic clone ratio, including the evolving genetic
767 aberrations, are shown for each case. (c) A comparison of the duration of the event-free time

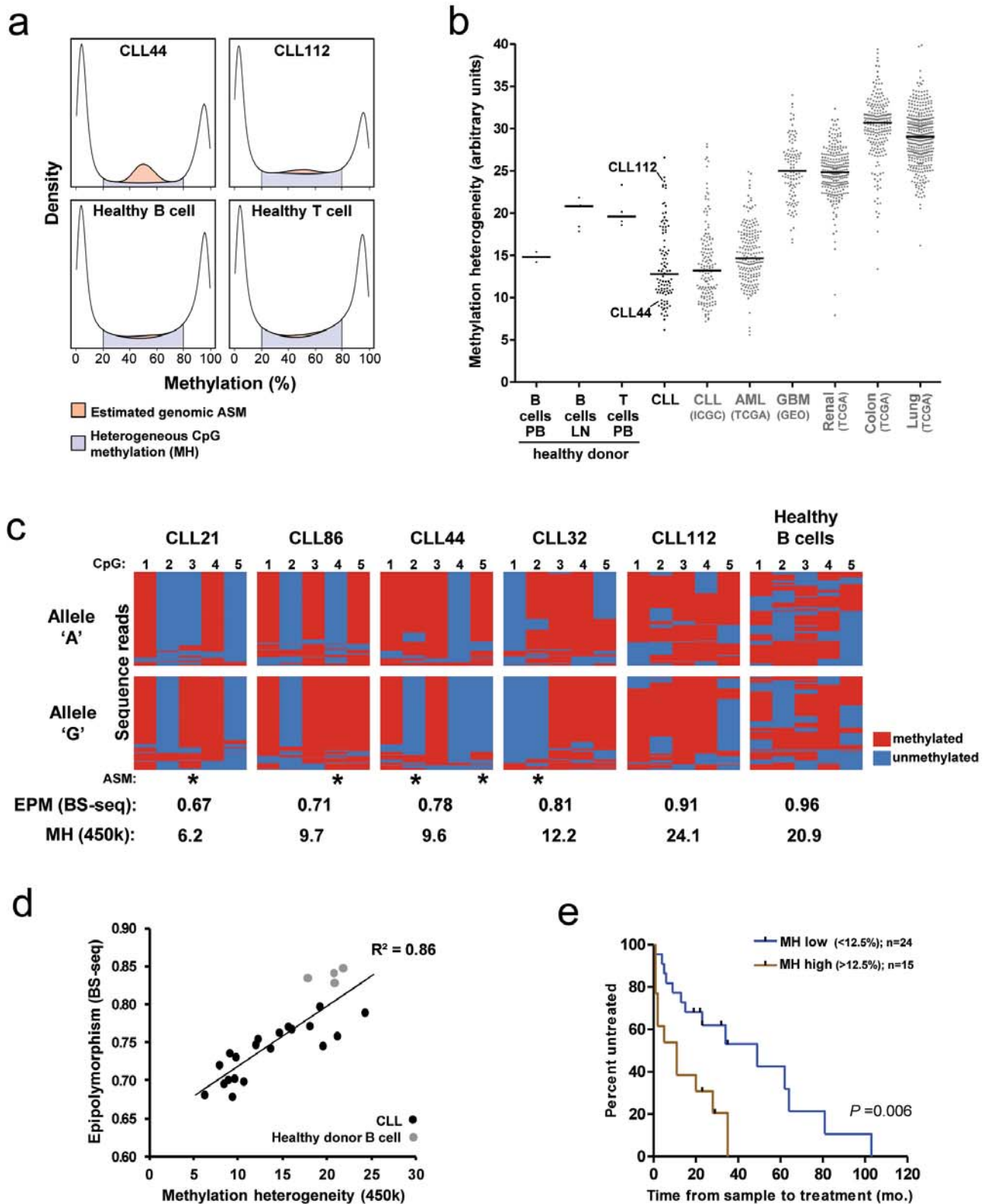
768 window following first-line therapy between CLL cases with high and no/low methylation
769 evolution. Second treatment or death was used as post-therapy events. Statistical analysis
770 performed by Mantel-Cox log rank test ($P < 0.0001$).

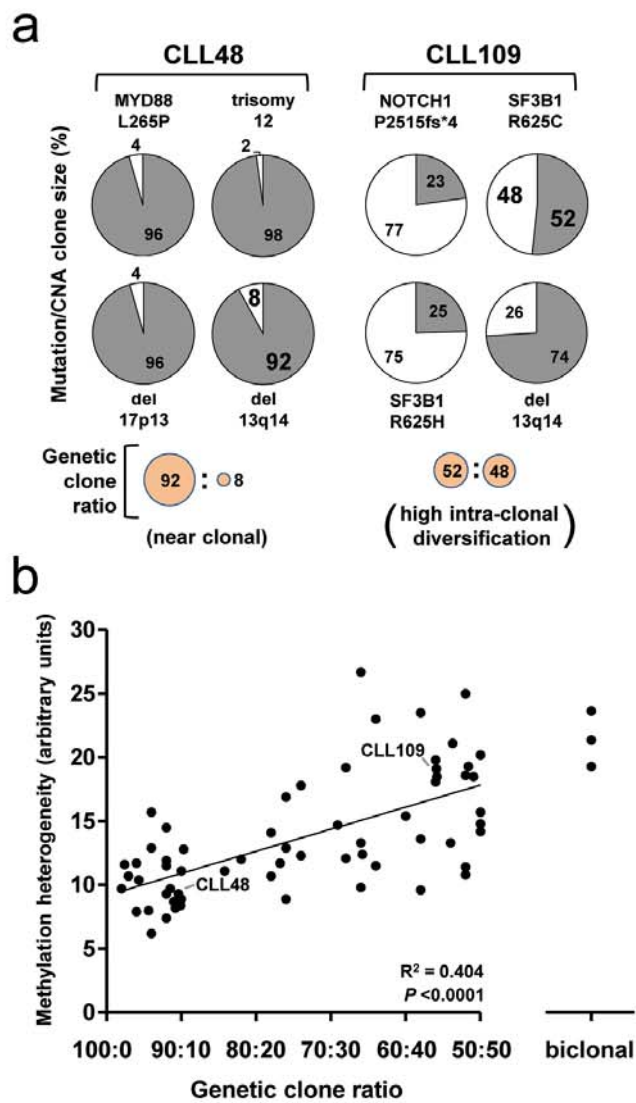
771

772 **Figure 6: Scenarios involving epigenetic and genetic evolution in the two-disease subtype**
773 **model of CLL.** CLL initiating events include genome-wide hypomethylation, which produces a
774 high degree of ASM, and usually a somatic genetic event, which together are observed as clonal
775 aberrations in all timepoints. In the *IGHV*-mutated subtype, the genome-wide DNA methylation
776 pattern of the founder cell is maintained with relative high fidelity. Selection of subclonal
777 populations with widespread epigenetic changes is not observed. Genetic evolution
778 independent of methylation evolution is only rarely observed and frequently involves a
779 recurrent deletion that includes 13q14. All cases that exhibit a high degree of methylation
780 evolution are the *IGHV*-unmutated disease subtype and involve simultaneous selection of
781 genetic aberrations. Two possible (non-mutually exclusive) hypotheses for coincident evolution
782 are shown: A) *simultaneous acquisition*, where the acquisition of a genetic subclonal driver
783 aberration directly impacts the epigenetic state of the subclonal founder cell, and B) *Step-wise*
784 *acquisition*, where a low-level of epigenetic stability precedes the acquisition of a genetic
785 subclonal driver, and thus a novel epigenetic pattern is co-selected with the genetic aberration.
786 SHM, somatic hypermutation.

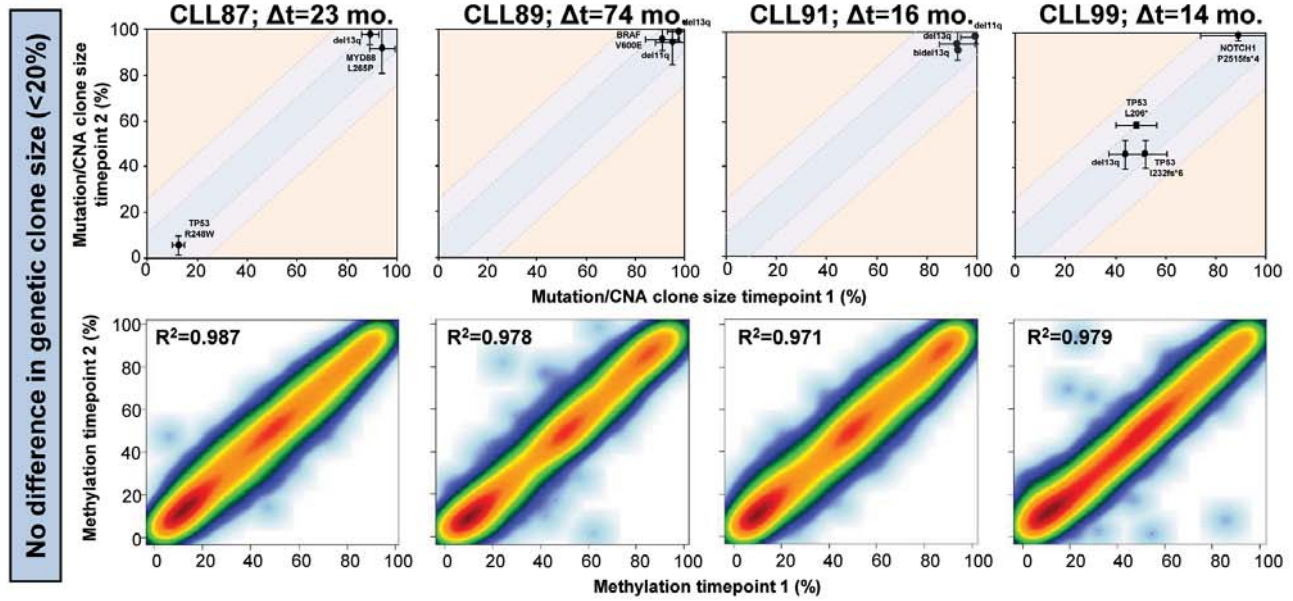
Figure 1:



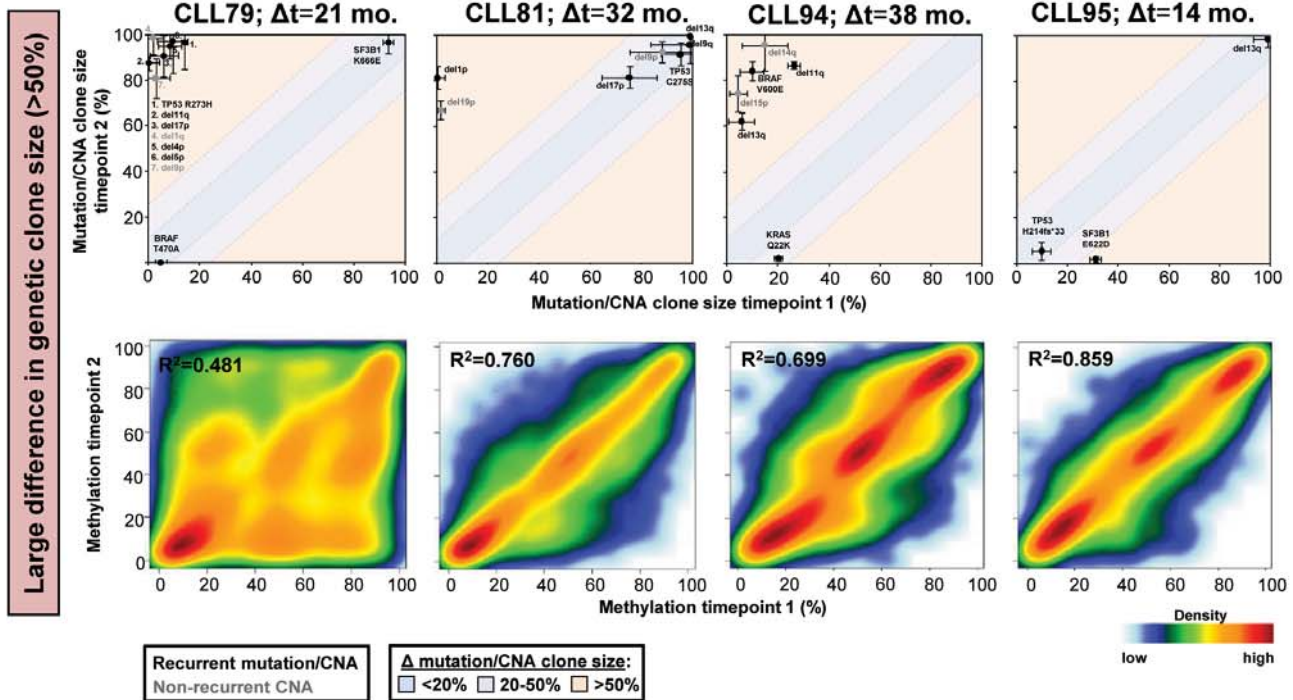


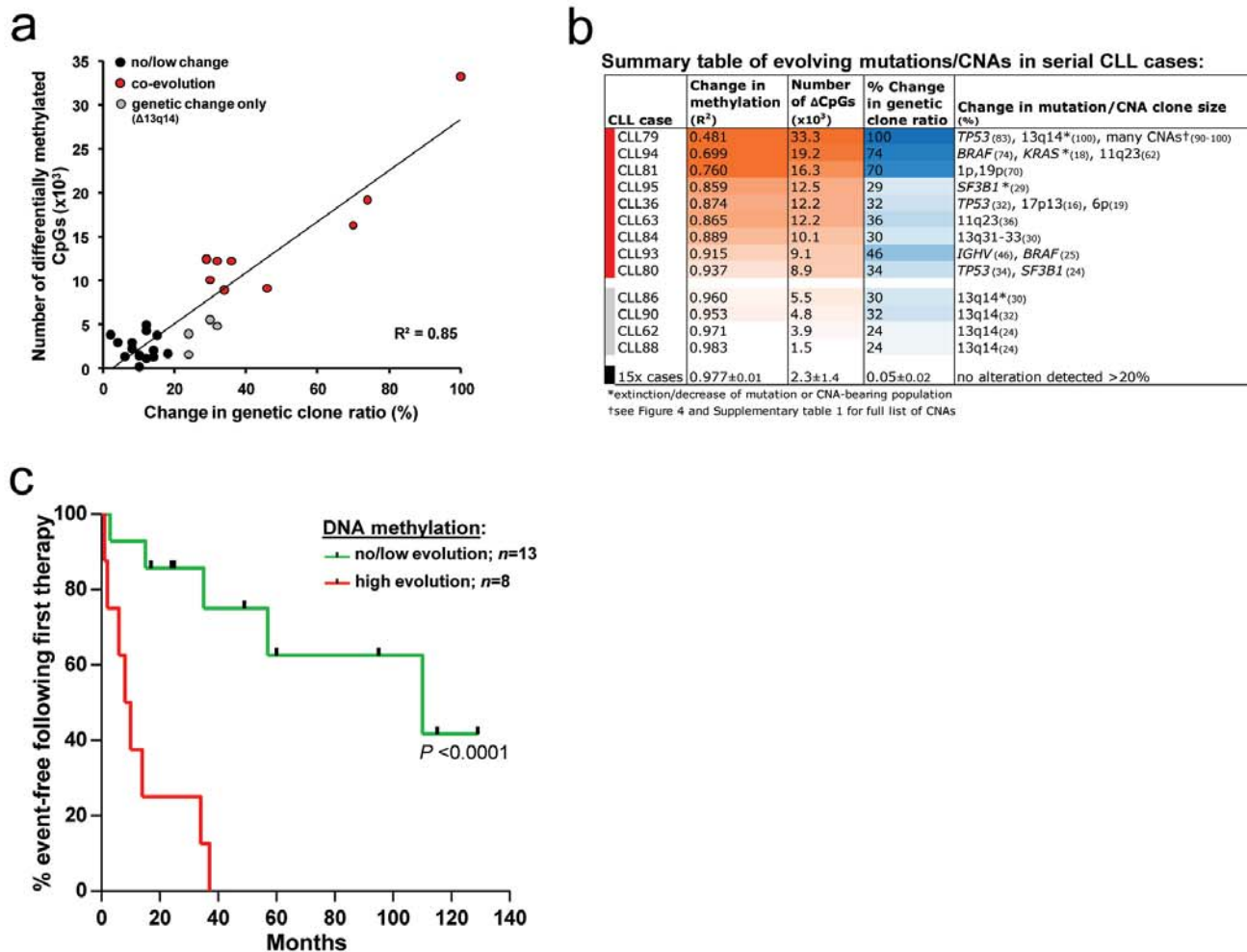


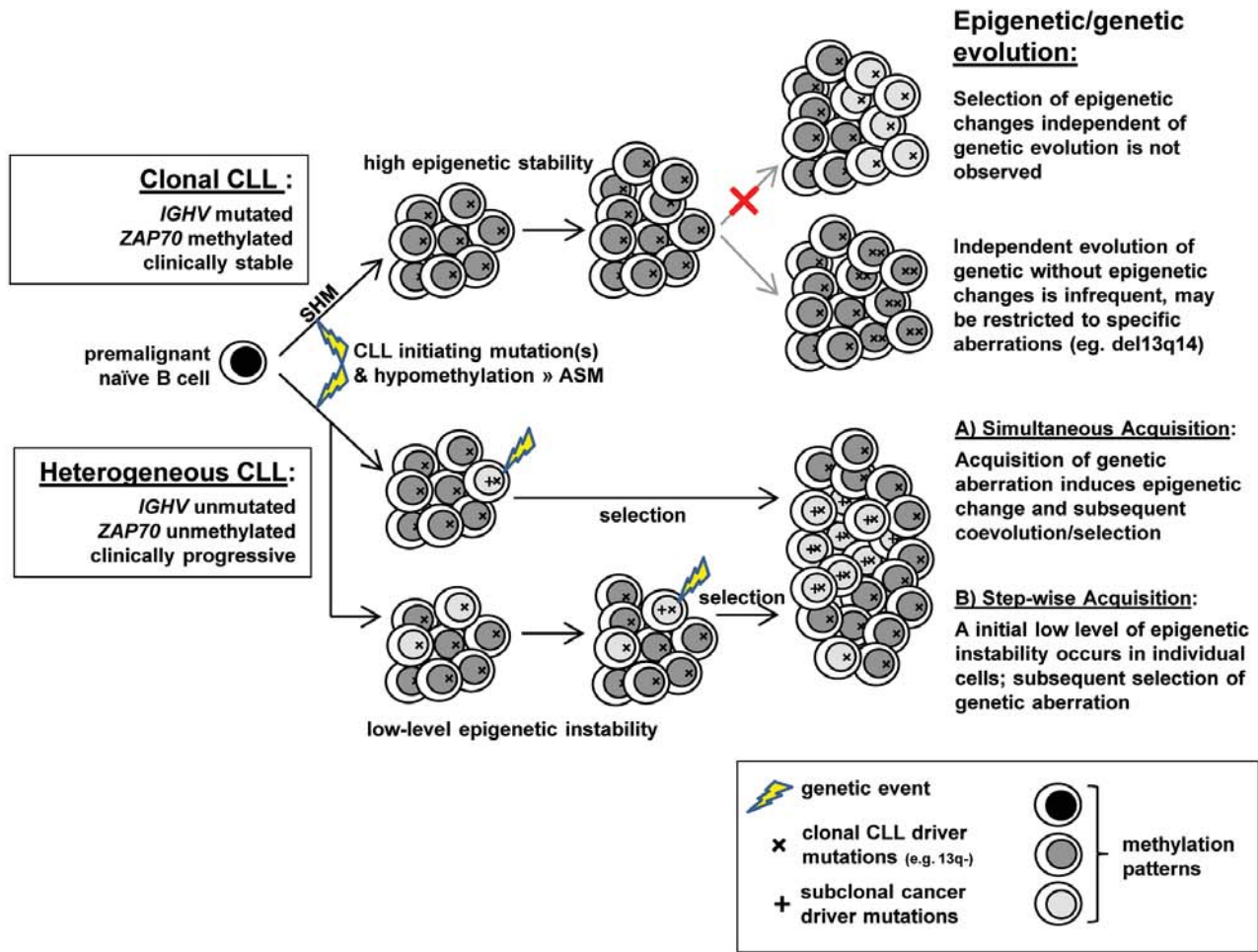
a



b







CANCER DISCOVERY

Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia

Christopher C Oakes, Rainer Claus, Lei Gu, et al.

Cancer Discovery Published OnlineFirst December 19, 2013.

Updated version	Access the most recent version of this article at: doi: 10.1158/2159-8290.CD-13-0349
Supplementary Material	Access the most recent supplemental material at: http://cancerdiscovery.aacrjournals.org/content/suppl/2013/12/18/2159-8290.CD-13-0349.DC1
Author Manuscript	Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link <http://cancerdiscovery.aacrjournals.org/content/early/2013/12/19/2159-8290.CD-13-0349>. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.