Author Manuscript Published OnlineFirst on December 19, 2013; DOI: 10.1158/2159-8290.CD-13-0349
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

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

Christopher C. Oakes[1], Rainer Claus[1,5], Lei Gu[1,2], Yassen Assenov[1], Jennifer Hüllein[6], Manuela Zucknick[3], Matthias Bieg[2], David Brocks[1], Olga Bogatyrova[1], Christopher R. Schmidt[1], Laura Rassenti[8], Thomas J. Kipps[8], Daniel Mertens[4,9], Peter Lichter[4,11], Hartmut Döhner[9], Stephan Stilgenbauer[9], John C. Byrd[10], Thorsten Zenz[6,7] and Christoph Plass[1,11]

Affiliations: Division of [1]Epigenomics and Cancer Risk Factors, [2]Theoretical Bioinformatics, [3]Biostatistics and [4]Molecular Genetics, The German Cancer Research Center (DKFZ), Heidelberg, Germany; [5]Department of Medicine, University of Freiburg Medical Center, Freiburg, Germany; [6]Department of Translational Oncology, National Center for Tumor Diseases (NCT) and German Cancer Research Center (DKFZ), Heidelberg, Germany; [7]Department of Medicine V, University of Heidelberg, Heidelberg, Germany; [8]Department of Medicine, University of California at San Diego Moores Cancer Center, La Jolla, USA; [9]Department of Internal Medicine III, University of Ulm, Ulm, Germany; [10]Division of Hematology, The Ohio State University, Columbus, USA; [11]The German Cancer Consortium.

Corresponding Author:

Christoph Plass

German Cancer Research Center (DKFZ),

Division of Epigenomics and Cancer Risk Factors,

Im Neuenheimer Feld 280, 69120 Heidelberg, Germany,

Tel. +49 6221-423330,

Fax +49 6221-423359,

E-mail: c.plass@dkfz.de

Running Title: Coevolution of epigenetics and genetics in CLL

Key words: DNA methylation; heterogeneity; coevolution; allele-specific methylation; chronic lymphocytic leukemia

Abbreviations list:

CLL, chronic lymphocytic leukemia

CNA, copy number alteration

ASM, allele-specific methylation

BS-seq, bisulfite sequencing

SNP, single nucleotide polymorphism

WGBS, whole-genome bisulfite sequencing

MH, methylation heterogeneity

EPM, epipolymorphism

AML, acute myeloid leukemia

NBC, naïve B cell

ncsMBC, non-class-switched memory B cell

csMBC, class-switched memory B cell

LN, lymph node

PB, peripheral blood

SHM, somatic hypermutation

Notes:

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

Conflict of interest disclosure statement:

The authors declare no competing financial interests.

ABSTRACT:

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

SIGNIFICANCE:

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

INTRODUCTION:

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

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

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

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

RESULTS:

CLL retains a large quantity of allele-specific methylation:

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

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

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

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

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

Methylation heterogeneity in CLL:

To investigate the basis for the high degree of variation in ASM between CLL samples, we hypothesized that the degree of ASM reflects intra-sample heterogeneity of DNA methylation patterns. In a diploid cell, CpG methylation values are restricted to three states (methylated, unmethylated and monoallelically methylated). If a population of cells maintains a stable, clonal pattern of methylation, values derived from a sample containing large numbers (usually $>1.0x10^7$) of cells will also be restricted to these three discrete ranges of CpG methylation values. As all CpGs found within CNAs have been removed from the analysis, methylation values that occur between these discrete ranges can only be caused by a disparate CpG methylation state between cells within the sample. The total amount of CpGs that fall outside the expected ranges can thus be used to estimate the overall level of methylation heterogeneity (MH) in a given sample.

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

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

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

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

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

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

The relationship between epigenetic and genetic heterogeneity:

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

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

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

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

In total, 13 of the 28 serial cases were observed to undergo a genetic change of >20% (Fig.5a,b). By defining a difference between groups by both the number of CpGs that differ by >10% and the R^2-value (see Supplemental figure 8 for a detailed description of group dichotomization), 9 of the 13 cases display evolution of methylation as defined by more than $5x10^3$ differentially methylated CpGs and $R^2<0.95$. The genetic aberrations that are observed to evolve co-dependently with methylation involve a subset of recurrent mutations and/or CNAs. The majority of these aberrations (i.e. those involving *TP53*, *SF3B1*, *BRAF*, del11q23, del17p13, etc.) have been previously described as subclonal cancer driver mutations that are frequently associated with genetic evolution (11). Decrease or extinction of some mutations/CNAs is found to occur, indicating that a potential hierarchy of aberrations exists between subclonal populations. The four cases that do not show evolution of methylation, yet show a >20% change in genetic clone ratio, specifically involve changes detected solely at the 13q14 locus, hinting that some aberrations may not be linked to methylation evolution. Epigenetic and genetic changes are highly co-dependent (Fisher's exact test, $P<0.001$), as widespread epigenetic evolution independent of genetic evolution is not observed.

Prediction and outcome of methylation evolution:

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

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

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

DISCUSSION:

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

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

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

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

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

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

MATERIALS AND METHODS:

CLL and healthy donor lymphocyte samples

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

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

Isolation and purification of CLL and healthy lymphocytes

All samples were obtained from whole blood then subjected to Ficoll-Isopaque density centrifugation, and CD19+ B and CD3+ T cells were isolated by positive magnetic cell separation (Miltenyi Biotec). Sorted cells were checked for purity by FACS with CD19/CD20 for healthy control samples and CD19/CD20/CD5 for CLL samples (BD Biosciences). Following sorting, all samples with a CD19/CD20/CD5 purity <98% were subjected to additional sorting, and the average final purity of all sorted samples was >99%. CLL samples with >$100x10^6$ WBC/ul were not subject to purification. DNA was extracted from purified cells using the Qiagen DNeasy kit (Qiagen) and quantified using a ND-100 spectrophotometer (Thermo Scientific).

DNA methylation analysis using 450k BeadChip arrays

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

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

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

Analysis of ASM using WGBS

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

Targeted bisulfite sequencing and analysis

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

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

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

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

REFERENCES:

1. Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. Nat Rev Cancer. 2011;11:726-34.

2. Broske AM, Vockentanz L, Kharazi S, Huska MR, Mancini E, Scheller M, et al. DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat Genet. 2009;41:1207-U69.

3. Trowbridge JJ, Snow JW, Kim J, Orkin SH. DNA Methyltransferase 1 Is Essential for and Uniquely Regulates Hematopoietic Stem and Progenitor Cells. Cell Stem Cell. 2009;5:442-9.

4. Timp W, Feinberg AP. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. Nat Rev Cancer. 2013.

5. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. The New England journal of medicine. 2012;366:883-92.

6. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. Cell. 2012;149:994-1007.

7. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. Nature. 2011;472:90-4.

8. Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA, et al. Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. Science. 2008;322:1377-80.

9. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012;481:506-10.

10. Schuh A, Becq J, Humphray S, Alexa A, Burns A, Clifford R, et al. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. Blood. 2012;120:4191-6.

11. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, et al. Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia. Cell. 2013;152:714-26.

12. Cahill N, Bergh AC, Kanduri M, Goransson-Kultima H, Mansouri L, Isaksson A, et al. 450K-array analysis of chronic lymphocytic leukemia cells reveals global DNA methylation to be relatively stable over time and similar in resting and proliferative compartments. Leukemia. 2013;27:150-8.

13. Stilgenbauer S, Sander S, Bullinger L, Benner A, Leupolt E, Winkler D, et al. Clonal evolution in chronic lymphocytic leukemia: acquisition of high-risk genomic aberrations associated with unmutated VH, resistance to therapy, and short survival. Haematol-Hematol J. 2007;92:1242-5.

14. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. Increased methylation variation in epigenetic domains across cancer types. Nat Genet. 2011;43:768-U77.

15. Landan G, Cohen NM, Mukamel Z, Bar A, Molchadsky A, Brosh R, et al. Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. Nat Genet. 2012;44:1207-14.

16. Rush LJ, Raval A, Funchain P, Johnson AJ, Smith L, Lucas DM, et al. Epigenetic profiling in chronic lymphocytic leukemia reveals novel methylation targets. Cancer Res. 2004;64:2424-33.

17. Kulis M, Heath S, Bibikova M, Queiros AC, Navarro A, Clot G, et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat Genet. 2012;44:1236-42.

18. Kanduri M, Cahill N, Goransson H, Enstrom C, Isaksson A, Rosenquist R. Differential Genome-Wide Methylation Profiles in Prognostic Subsets of Chronic Lymphocytic Leukemia. Haematol-Hematol J. 2009;94:200-.

19. Claus R, Lucas DM, Stilgenbauer S, Ruppert AS, Yu LB, Zucknick M, et al. Quantitative DNA Methylation Analysis Identifies a Single CpG Dinucleotide Important for ZAP-70 Expression and Predictive of Prognosis in Chronic Lymphocytic Leukemia. J Clin Oncol. 2012;30:2483-91.

20. Ehrlich M, Lacey M. DNA hypomethylation and hemimethylation in cancer. Advances in experimental medicine and biology. 2013;754:31-56.

21. Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, Pelizzola M, et al. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome research. 2012;22:246-58.

22. Fang F, Hodges E, Molaro A, Dean M, Hannon GJ, Smith AD. Genomic landscape of human allele-specific DNA methylation. P Natl Acad Sci USA. 2012;109:7332-7.

23. Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J, et al. Base-Resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. Cell. 2012;148:816-31.

24. Sturm D, Witt H, Hovestadt V, Khuong-Quang DA, Jones DTW, Konermann C, et al. Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma. Cancer Cell. 2012;22:425-37.

25. Network CGAR. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. The New England journal of medicine. 2013;368:2059-74.

643 26. TCGA. The Cancer Genome Atlas Data Portal. https://tcga-
644 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. Compehensive 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. Edelmann 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. (%)**	***P*-value**
Patient characteristics:			
Age at diagnosis (yr. ±s.d.)	55.8 ±11.8	59.9 ±10.6	*n.s.*
Sex (Female)	10 (33)	15 (48)	*n.s.*
Pre-treatment (Yes)	6 (19)	13 (42)	**0.014**
Prognostic indicators:			
IGHV unmutated	8 (35)	11 (79)	**<0.01**
ZAP70 low methylation	7 (30)	9 (69)	**0.024**
Cytogenetics (NK, sole -13q)	14 (61)	9 (69)	*n.s.*

695

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

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

697

FIGURE LEGENDS:

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

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

Figure 3: **Epigenetic heterogeneity is associated with genetic heterogeneity in CLL samples.** (a) Quantitative assessment of the mutation/CNA clone sizes for various aberrations in two CLL samples representative of different levels of genetic heterogeneity. Clone sizes for various detected somatic aberrations (grey) are displayed for CLL48 and CLL109. For CLL48, all variations fall within a range consistent with a clonal sample population possessing monoallelic aberrations at 13q14, 17p13 and *MYD88* and a single-copy gain of chromosome 12. For CLL109, individual somatic variations occur at a frequency indicative of intraclonal diversification, with the mutation clone size of the R625C mutation in *SF3B1* representing approximately an even ratio of 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 (Δ<20%) and (b) large change (Δ>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%, 20-50% 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 Δ>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

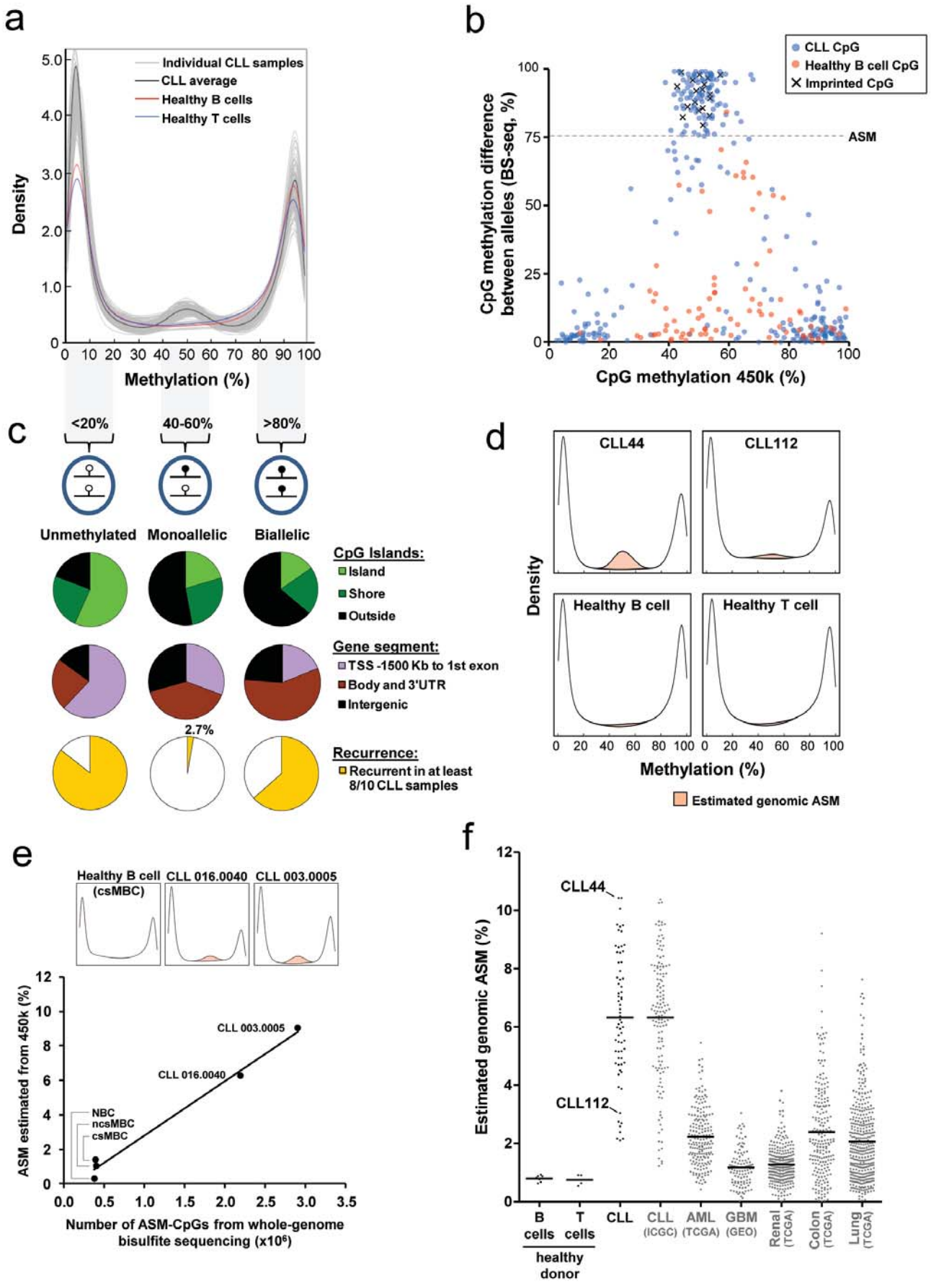
768 window following first-line therapy between CLL cases with high and no/low methylation
769 evolution. Second treatment or death was used as post-therapy events. Statistical analysis
770 performed by Mantel-Cox log rank test (P<0.0001).

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.

Author Manuscript Published OnlineFirst on December 19, 2013; DOI: 10.1158/2159-8290.CD-13-0349
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Figure 1:



Author Manuscript Published OnlineFirst on December 19, 2013; DOI: 10.1158/2159-8290.CD-13-0349
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Figure 2:

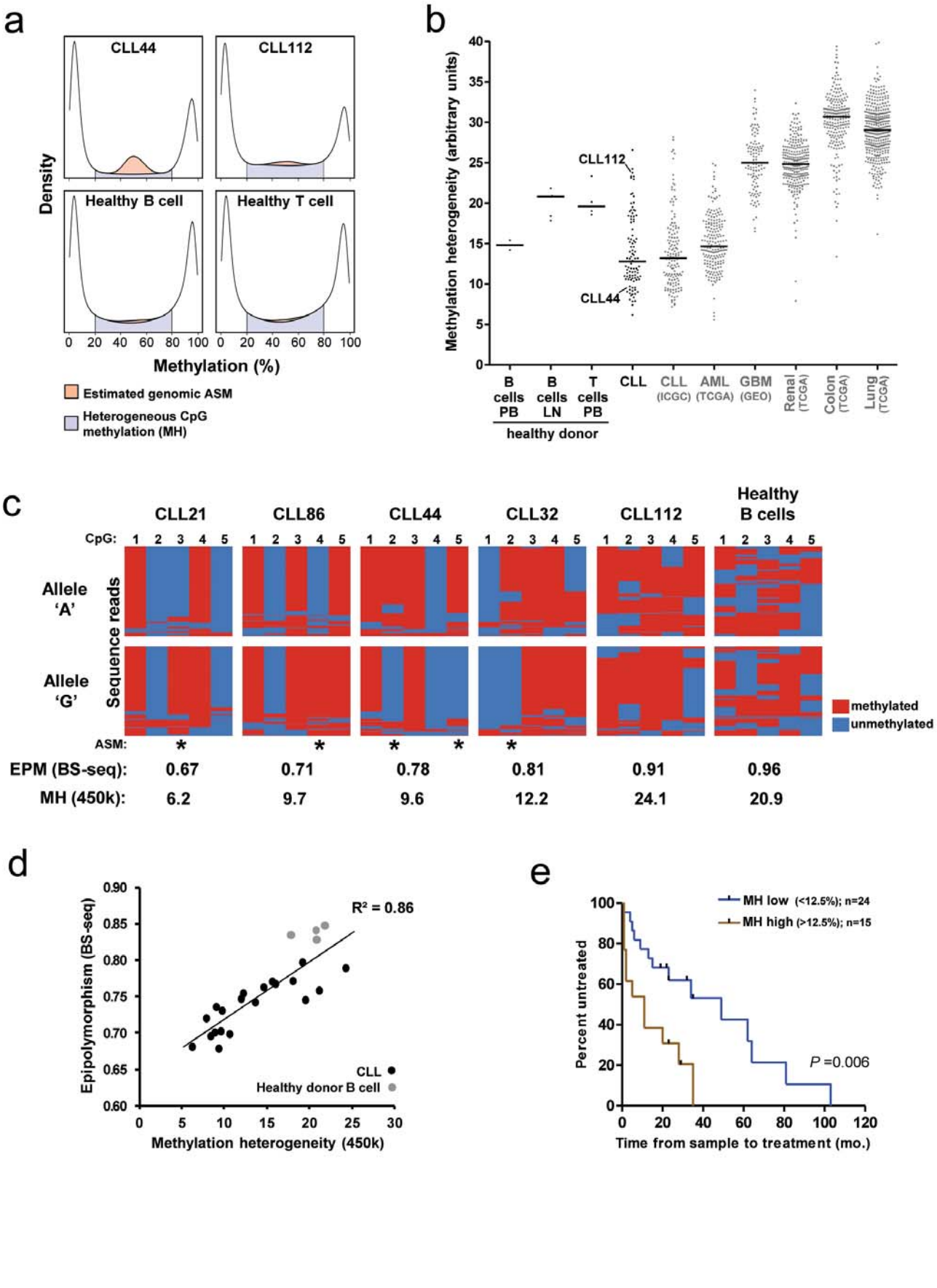
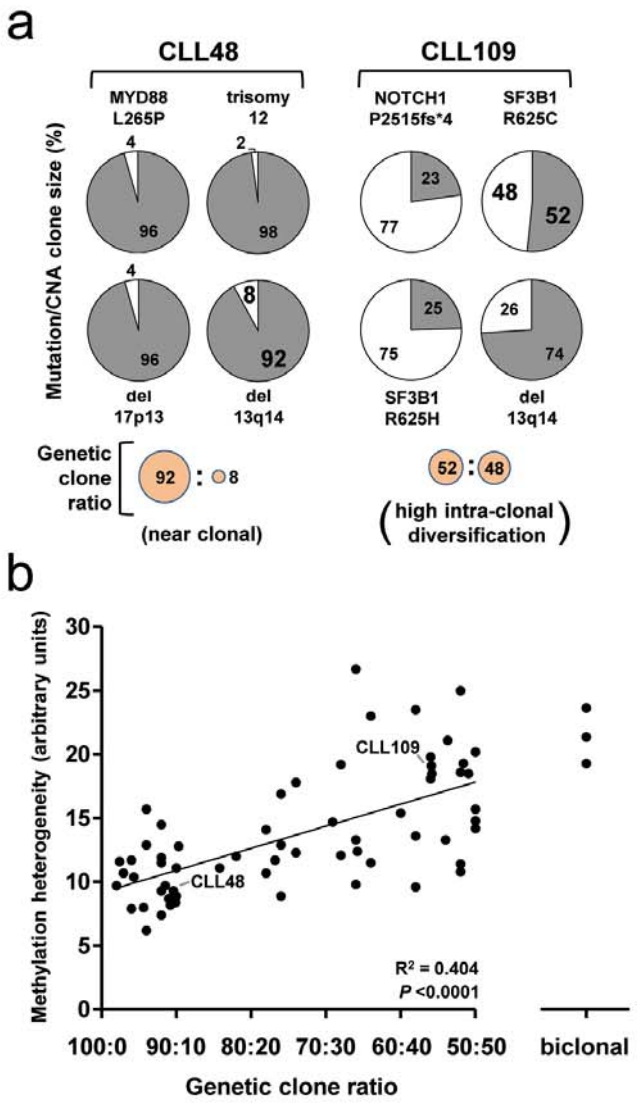


Figure 3: Author Manuscript Published OnlineFirst on December 19, 2013; DOI: 10.1158/2159-8290.CD-13-0349
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.



Author Manuscript Published OnlineFirst on December 19, 2013; DOI: 10.1158/2159-8290.CD-13-0349
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Figure 4:

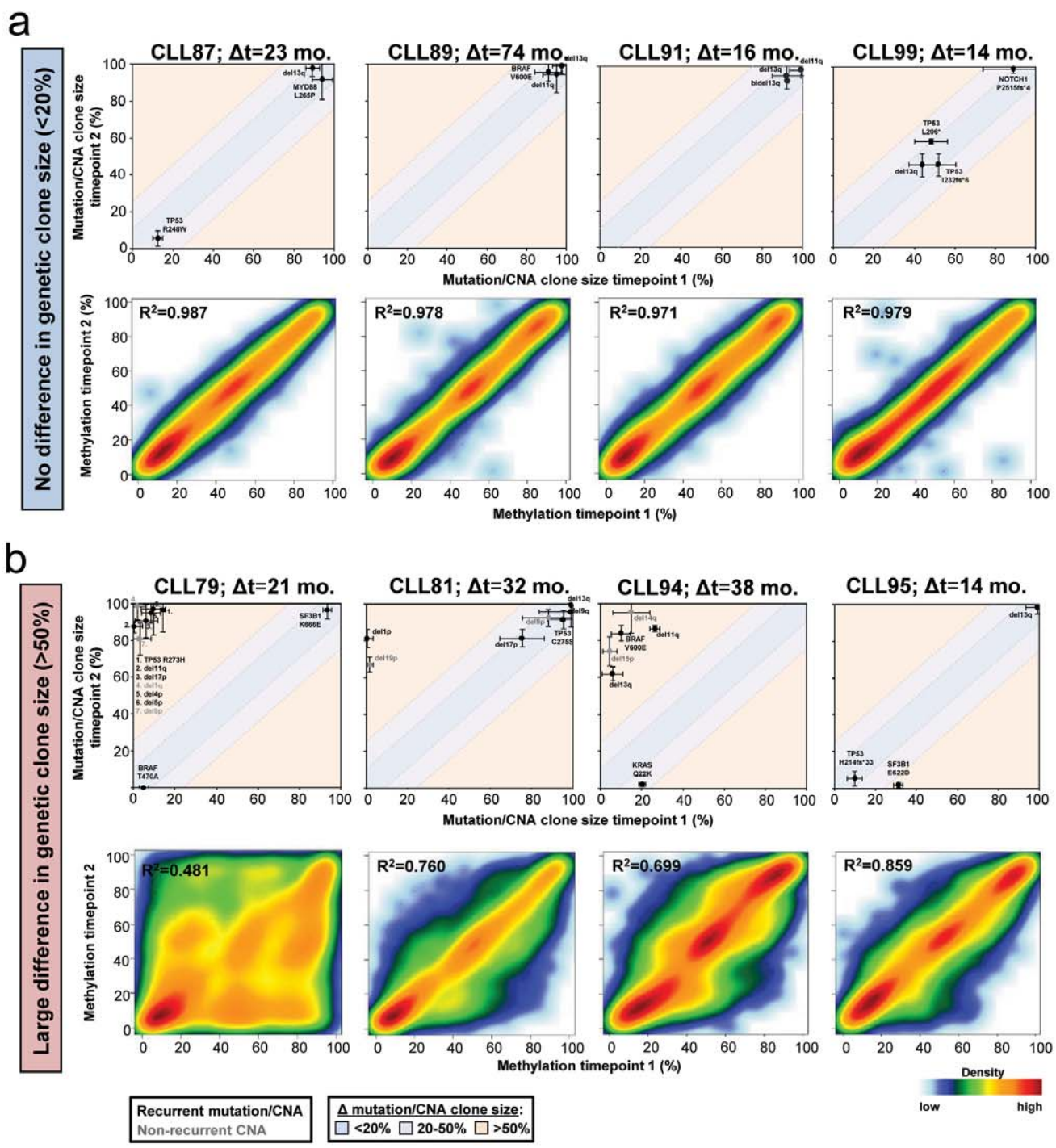
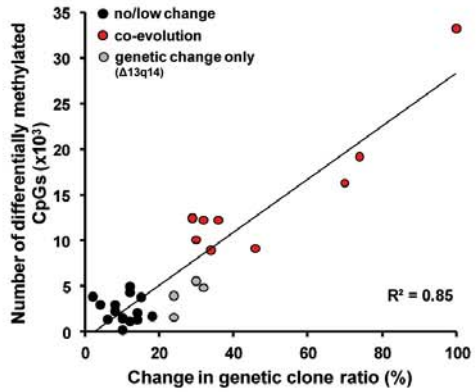


Figure 5: Author Manuscript Published OnlineFirst on December 19, 2013; DOI: 10.1158/2159-8290.CD-13-0349
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

a



b

Summary table of evolving mutations/CNAs in serial CLL cases:

CLL case	Change in methylation (R^2)	Number of ΔCpGs ($\times 10^3$)	% Change in genetic clone ratio	Change in mutation/CNA clone size (%)
CLL79	0.481	33.3	100	*TP53* (83), 13q14* (100), many CNAs† (90-100)
CLL94	0.699	19.2	74	*BRAF* (74), *KRAS* * (18), 11q23 (62)
CLL81	0.760	16.3	70	1p,19p (70)
CLL95	0.859	12.5	29	*SF3B1* * (29)
CLL36	0.874	12.2	32	*TP53* (32), 17p13 (16), 6p (19)
CLL63	0.865	12.2	36	11q23 (36)
CLL84	0.889	10.1	30	13q31-33 (30)
CLL93	0.915	9.1	46	*IGHV* (46), *BRAF* (25)
CLL80	0.937	8.9	34	*TP53* (34), *SF3B1* (24)
CLL86	0.960	5.5	30	13q14* (30)
CLL90	0.953	4.8	32	13q14 (32)
CLL62	0.971	3.9	24	13q14 (24)
CLL88	0.983	1.5	24	13q14 (24)
15x cases	0.977±0.01	2.3±1.4	0.05±0.02	no alteration detected >20%

*extinction/decrease of mutation or CNA-bearing population
†see Figure 4 and Supplementary table 1 for full list of CNAs

c

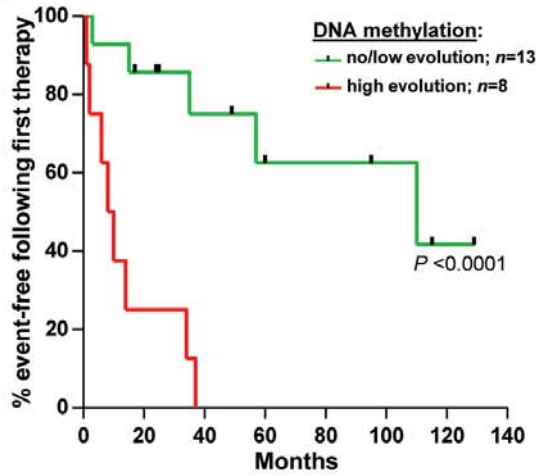
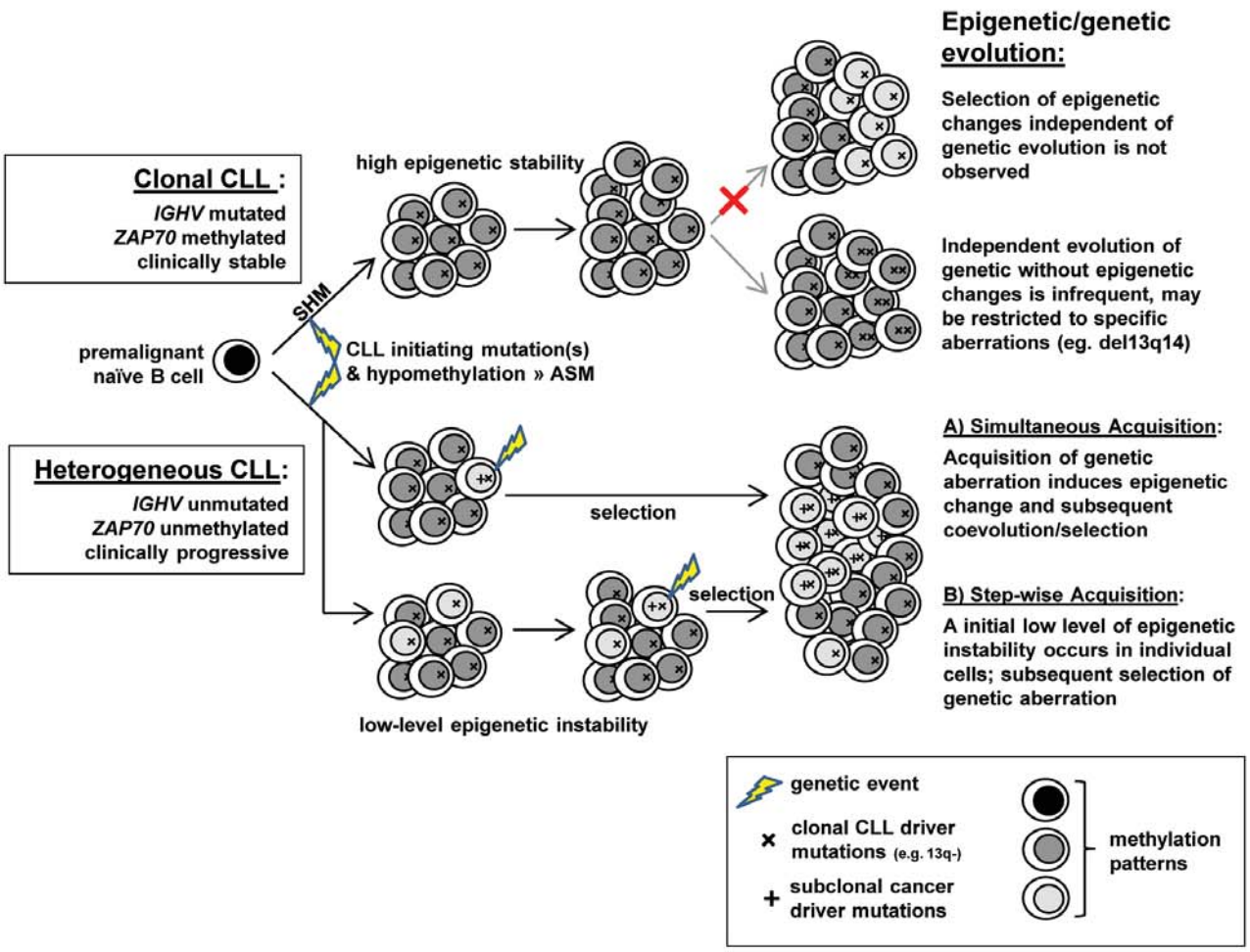


Figure 6: Author Manuscript Published OnlineFirst on December 19, 2013; DOI: 10.1158/2159-8290.CD-13-0349
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.



AACR American Association for Cancer Research

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.