Analysis

The proliferative history shapes the DNA methylome of B-cell tumors and predicts clinical outcome

Martí Duran-Ferrer[1,2,*], Guillem Clot[1,2], Ferran Nadeu[1,2], Renée Beekman[1,2], Tycho Baumann[2,3], Jessica Nordlund[4], Yanara Marincevic-Zuniga[4], Gudmar Lönnerholm[5], Alfredo Rivas-Delgado[1,3], Silvia Martin[1,2], Raquel Ordoñez[2,6], Giancarlo Castellano[1], Marta Kulis[1], Ana Queirós[1], Lee Seung-Tae[7], Joseph Wiemels[8], Romina Royo[9], Montserrat Puiggrós[9], Junyan Lu[10], Eva Gine[1,2,3], Sílvia Beà[1,2,13], Pedro Jares[1,2,13], Xabier Agirre[2,6], Felipe Prosper[2,6,11], Carlos López-Otín[2,12], Xosé S.Puente[2,12], Christopher C. Oakes[13], Thorsten Zenz[13,14,], Julio Delgado[1,2,3], Armando López-Guillermo[1,2,3], Elías Campo[1,2,15] and José Ignacio Martin-Subero[1,2,15,16*]

[1]Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain.
[2]Centro de Investigación Biomédica en Red de Cáncer, CIBERONC, Spain.
[3]Servicio de Hematología, Hospital Clínic, IDIBAPS, Barcelona, Spain.
[4]Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden
[5]Department of Women's and Children's Health, Pediatrics, Uppsala University, Uppsala, Sweden.
[6]Centro de Investigación Médica Aplicada (CIMA), IDISNA, Pamplona, Spain.
[7]Department of Laboratory Medicine, Yonsei University College of Medicine, Korea.
[8]Center for Genetic Epidemiology, University of Southern California, Los Angeles.
[9]Programa Conjunto de Biología Computacional, Barcelona Supercomputing Center (BSC), Institut de Recerca Biomèdica (IRB), Spanish National Bioinformatics Institute, Universitat de Barcelona, Barcelona, Spain.
[10]European Molecular Biology Laboratory (EMBL), Heidelberg, Germany.
[11]Hematology and Cell Therapy Department, Clínica Universidad de Navarra, Universidad de Navarra, Avenida Pío XII, 36 31008 Pamplona, Spain.
[12]Departamento de Bioquímica y Biología Molecular, Instituto Universitario de Oncología (IUOPA), Universidad de Oviedo, Oviedo, Spain.
[13]Division of Hematology, Department of Internal Medicine, The Ohio State University, Columbus, OH
[14]Department of Medical Oncology and Hematology, University Hospital Zürich and University of Zürich, Zürich, Switzerland
[15]Departament de Fonaments Clinics, Facultat de Medicina, Universitat de Barcelona, Barcelona, Spain
[16]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

[*]Correspondence: maduran@clinic.cat, imartins@clinic.cat

Abstract

We report a systematic analysis of the biological and clinical implications of DNA methylation variability in five categories of B-cell tumors derived from B cells spanning the entire maturation spectrum. We used 2056 primary samples including training and validation series and show that 88% of the human DNA methylome is dynamically modulated under normal and neoplastic conditions. B-cell tumors display both epigenetic imprints of their cellular origin and *de novo*, disease-specific epigenetic alterations that in part are related to differential transcription factor binding. These differential methylation patterns were used by a machine-learning approach to create a diagnostic algorithm that accurately classifies 14 B-cell tumor entities and subtypes with different clinical management. Beyond this, we identified extensive patient-specific epigenetic variability targeting constitutively silenced chromatin regions, a phenomenon we could relate to the proliferative history of normal and neoplastic B cells. We observed that, depending on the maturation stage of the tumor cell of origin, mitotic activity leaves different imprints into the DNA methylome. Subsequently, we constructed a novel DNA methylation-based mitotic clock called epiCMIT (epigenetically-determined Cumulative MIToses), whose lapse magnitude represents a strong independent prognostic variable within specific B-cell tumor subtypes and is associated with particular driver genetic alterations. Our findings reveal DNA methylation as a holistic tracker of B-cell tumor developmental history, with implications in the differential diagnosis and prediction of the outcome of the patients.

67 **Introduction**

68 The process of neoplastic transformation implies a dramatic alteration of cellular
69 identity [1]. However, cancer cells partially maintain molecular imprints of the cellular
70 lineage and maturation stage from which they originate [2]. B-cell neoplasms are a
71 paradigmatic model of this phenomenon, as the maturation stage of different B-cell
72 neoplasms is the main principle behind the World Health Organization classification of
73 these tumors [3]. Recent studies have focused on the analysis of the DNA methylome, a
74 *bona fide* epigenetic mark related to cellular identity and gene regulation [4,5], during
75 the entire B-cell maturation program [6] and in various B-cell neoplasms spanning the
76 entire maturation spectrum. These include B-cell acute lymphoblastic leukemia (ALL)
77 [7,8] derived from precursor B cells, mantle cell lymphoma (MCL) [9] and chronic
78 lymphocytic leukemia [10,11](CLL) derived from pre- and post-germinal center mature B
79 cells, diffuse large B-cell lymphoma (DLBCL) [12,13] derived from germinal center B cells,
80 and multiple myeloma (MM) [14,15] derived from terminally-differentiated plasma cells.
81 These studies have revealed a dynamic DNA methylome during B-cell maturation as
82 well as novel insights into the cellular origin, pathogenic mechanisms and clinical
83 behavior of B-cell neoplasms reviewed in [16]. However, a global analysis of the entire
84 normal cell differentiation program and derived neoplasms is neither available for B
85 cells nor for any other human cell lineage. Thus, we herein exploit both previously
86 generated DNA methylation datasets as well newly generated data to systematically
87 decipher the sources of DNA methylation variability across B-cell neoplasms. This
88 comprehensive approach using over 2000 samples reveals previously hidden biological
89 insights and clinical associations. In particular, *de novo* disease-specific
90 hypomethylation in active regulatory regions is associated with differential

91 transcription factor binding and targets genes important for disease-specific
92 pathogenesis. From the clinical perspective, we define a set of epigenetic biomarkers
93 that can accurately classify B-cell neoplasms requiring differential clinical management
94 and construct a DNA methylation-based mitotic clock called epiCMIT as a personalized
95 predictor of clinical behavior within each B-cell neoplasm.

96

97 ## Results

98 **Shared DNA methylation dynamics in normal and neoplastic B cells**

99 We analyzed previously published DNA methylation profiles of samples from normal
100 and neoplastic B cells spanning the entire B-cell differentiation spectrum, all generated
101 with the 450k microarray platform from Illumina. These included 10 normal B cell
102 subpopulations [6] as well as the five common categories of B-cell neoplasms, i.e. ALL [7,8],
103 MCL [9], CLL [10,17], DLBCL (own unpublished series) and MM [14] (Fig. 1a and Supplementary
104 Table 1). Following the guidelines of the TCGA Consortium, we selected samples
105 containing a tumor-cell content more than 60% https://www.cancer.gov/about-
106 nci/organization/ccg/blog/2018/bcr-tips. This proportion of tumor cells was estimated
107 by flow cytometry [6,9,10,14,17], genetic data [18] and/or lineage-specific DNA methylation
108 patterns (Supplementary Table 2). To validate that 60% was an appropriate threshold
109 for B-cell tumors, we analyzed 5 MCLs and 3 CLLs in which we had DNA extracted both
110 from sorted (minimum purity of 86%) and unsorted tumor cells (purity between 48%
111 and 77%). Unsupervised analyses showed that samples tightly clustered according to
112 patient number and not based on their tumor cell content (Extended Data Fig. 1a). In
113 fact, tumor cell content in these paired samples was reflected in a minor component
114 accounting for only 2% of the total variability, and therefore was considered negligible

115 (data not shown). Tumor cell content estimated by any of the three methods was in
116 general highly correlated (Extended Data Fig. 1b) and in most of the cases, we used
117 DNA methylation data to estimate tumor purity. However, MM samples showed that
118 DNA methylation-based estimation of tumor cell content was far lower than that
119 estimated by flow cytometry (Extended Data Fig. 1c). This result is congruent with
120 previous DNA methylation data indicating that MM losses the B cell identity [14].
121 Unexpectedly, some DLBCL samples also showed a similar effect (Extended Fig. 1c),
122 and therefore in MM and DLBCL, tumor cell content was estimated by flow cytometry
123 and genetic data, respectively. After all filtering criteria (Methods), we generated a
124 curated data matrix containing 1595 high quality samples (Fig. 1a and Supplementary
125 Table 1) with DNA methylation values for 437182 CpGs, which was used in all
126 downstream analyses.

127 This large comprehensive dataset offered an exceptional opportunity to step-
128 wise dissect the DNA methylation variability of normal and neoplastic B cells at
129 different magnitude levels, including cancer-specific, tumor entity-specific, tumor
130 subentity-specific and individual-specific variability (Fig. 1b). Before studying such DNA
131 methylation dynamics, we identified that only 12% of the studied CpGs show stable
132 DNA methylation levels in normal and neoplastic B cells (Fig. 1C and Supplementary
133 Table 3). We characterized these CpGs based on genetic location, CpG content and
134 chromatin states from primary samples from the Blueprint consortium [19]. We
135 identified that stably methylated regions were located at gene-bodies of actively
136 transcribed genes whereas stably unmethylated sites were prevalent in CpG islands of
137 promoter regions showing active chromatin marks (Fig. 1d, e). Stably methylated and
138 unmethylated CpGs mainly converged into the same genes (Fig. 1e and Extended Data

139 Fig. 1e), which are highly expressed in normal and neoplastic B cells (Fig. 1f) and are
140 involved in cellular functions such as cell cycle, RNA processing and energy metabolism
141 (Extended Data Fig. 1g). These results indicate, on the one hand, that the stable
142 fraction of the DNA methylome affects genes involved in fundamental cellular
143 functions showing an epigenetic signature of expressed genes (Extended Data Fig. 1g).
144 On the other hand, they show that the great majority of the DNA methylome (88%) is
145 labile during normal B cell differentiation and neoplastic transformation.

146 We next wondered whether all B-cell tumors, regardless the entity, share a
147 unifying DNA methylation signature related to their neoplastic nature. As 25% of the
148 DNA methylome is modulated during normal B cell differentiation [6,11], we took the
149 remaining stable 75% and compared the methylomes of normal and neoplastic B cells.
150 This analysis did not identify any consistent *de novo* DNA methylation signature that is
151 shared by all B-cell neoplasms under study. Instead, DNA methylation variability is
152 related to differences among B-cell tumor entities and subtypes as well as patient-
153 specific variability, as will be shown in the following sections.

154

155 **Disease-specific hypomethylation targeting regulatory regions is associated with**
156 **specific transcription factor bindings and differential gene expression**

157 We next aimed at studying whether disease-specific DNA methylation patterns may
158 unravel differential pathogenic mechanisms underlying each B-cell neoplasm. Overall,
159 a principal component analysis (PCA) showed that different B-cell neoplasms form
160 distinct clusters (Fig. 2a). A further analysis indicated that the first nine components
161 contained information regarding specific B-cell neoplasms and allowed us to distil the
162 main biological sources of DNA methylation variability (Extended Data Fig 2a). The first

163 component was related to B-cell development, separating neoplasms according to the
164 maturation stage of their cellular origin, i.e. ALL together with pre-germinal center B
165 cells and mature B cell neoplasms together with germinal center B cells, memory B
166 cells and plasma cells. The remaining components showed tumor-specific patterns,
167 such as PC3, PC7 and PC9 related to MM, DLBCL and MCL-specific variability,
168 respectively. Next, in order to identify DNA methylation signatures specifically
169 associated with malignant transformation, we focused our next analysis on the
170 genome fraction with stable (i.e. B-cell independent) DNA methylation levels across B-
171 cell differentiation [6] (Fig. 2b). We showed varying numbers of tumor-specific DNA
172 methylation (tsDNAm) changes, ranging from 616 in CLL to 49279 in MM (Fig. 2b,
173 Supplementary Tables 4 and 5). Remarkably, we observed that DNA methylation
174 changes manifested differently in distinct neoplasms. Overall, hypermethylation was
175 enriched at CpG islands and promoter related regions, whereas hypomethylation at
176 low CpG content regions such as open sea, shore and shelfs (Extended Data Fig. 2c).
177 ALL and DLBCL showed more tumor-specific DNA hypermethylation (tsDNAm-hyper)
178 whereas MCL, CLL and MM acquired more tumor-specific DNA hypomethylation
179 (tsDNAm-hypo), being this skew towards hypomethylation remarkable in MM (Fig. 2b-
180 c and Extended Data Fig. 2b). To shed light into the potential causes of this
181 phenomenon, we analyzed the expression levels of DNA methyltransferases (DNMTs)
182 during normal B-cell maturation [6]. We identified that DNMT1, which has been linked
183 both to DNA methylation maintenance and DNA hypermethylation of CpG islands
184 marked by polycomb [20], shows the highest expression levels in precursor B cells and
185 germinal center B cells, the respective cells of origin of the mostly hypermethylated
186 ALL and DLBCL. Conversely, bone marrow plasma cells, the cell of origin of the mostly

187 hypomethylated MM, showed the lowest levels of DNMT1 expression (Extended Data
188 Fig. 2d). Although the precise mechanisms remain to be elucidated, the level of
189 expression of DNMTs of the respective cells of origin may influence how DNA
190 methylation changes are manifested in distinct B-cell neoplasms.

191 Transcription factors (TFs) have been reported to affect DNA methylation levels
192 of regulatory regions upon their binding to the DNA [21,22]. Therefore, we performed TFs
193 binding site prediction analysis in active regulatory elements (i.e. marked by H3K27ac)
194 containing tsDNAm-hypo CpG (Fig. 2d and Methods). Interestingly, the entities in
195 which tsDNAm-hypo was predominantly located in H3K27ac regions (Fig. 2c) showed
196 enrichments for binding sites of TFs expressed in each respective entity (Extended
197 Data Fig. 2e and Supplementary Table 6) and with a previously reported implication in
198 their pathogenesis, such as SPI1/SPIB and EBF1 in ALL, TCF/ZEB in MCL, and NFAT in
199 CLL (Fig. 2d) [23–25]. In the case of DLBCL and MM, their associated tsDNAm-hypo CpGs
200 were actually depleted of regulatory elements (Fig. 2c), suggesting that TF binding may
201 not be a major factor leading to their tumor-specific DNA methylation signatures.
202 However, focusing only on the underrepresented H3K27ac-containing tsDNAm-hypo
203 CpGs, we could also detect significant relationships with TFs potentially involved in the
204 respective diseases, such as FOX family in DLBCL [26], and NRL (a member of the
205 oncogenic MAF family), ISL1, TEAD, and YY1 in MM [27–30].

206 Beyond the potential role of TFs in shaping tumor-specific DNA methylation
207 signatures, we also investigated the downstream transcriptional consequences of
208 tsDNAm-hypo signatures. An analysis of transcriptional profiles of cases from all five
209 diseases revealed a total of 94 genes associated with tsDNAm-hypo genes expressed in
210 a disease-specific manner (Fig. 2e). Although some of the identified genes have been

211 shown to be specifically expressed in a particular disease, such as *CTLA4* and *KSR2* in
212 CLL [31], this comprehensive analysis provides a rich resource of novel epigenetically-
213 regulated candidate oncogenes for mechanistic studies in each B-cell neoplasm entity.
214

215 **Accurate classification of 14 clinico-biological subtypes of B cell neoplasms using**
216 **epigenetic biomarkers**

217 The B-cell neoplasms shown in Fig. 1a represent broad categories which are further
218 classified into subtypes with different clinico-biological features based on genetic,
219 transcriptional or epigenetic features [3]. These include high-hyperdiploid (HeH),
220 11q23/MLL, t(12;21), t(1;19), t(9;22) and dic(9;20) ALLs [7]; C1 (conventional; germinal
221 center-inexperienced) and C2 (leukemic non-nodal; germinal center-experienced)
222 MCLs [9]; naïve-like/low-programmed, intermediate/intermediate-programmed and
223 memory-like/high-programmed CLLs [10,11], and finally germinal center B cell (GCB) and
224 activated B cell (ABC) DLBCLs [32]. In MM, a previous report did not show methylation
225 differences among the distinct cytogenetic subtypes [14] and thus MM subgrouping was
226 not included in our analyses. Here, we focused on the identification of epigenetic
227 biomarkers that may allow a comprehensive diagnosis of B-cell tumors entities and
228 subtypes. We devised a strategy to construct a classifier algorithm that yielded 56
229 CpGs as the optimal number distributed along 5 predictors (Extended Data Fig. 3a-e
230 and Supplementary Table 7, Methods) to accurately discriminate the main B-cell tumor
231 entities as a first step (predictor 1), and subsequently B-cell tumor subtypes as a
232 second step (predictors 2, 3, 4 or 5) (Fig. 3a and Methods). The accuracy of the five
233 predictors was evaluated using nested 10-fold stratified cross-validation in the training
234 series (n=1345) and with external validation series (n=711) (Fig. 3b). Overall, we

235 obtained very high accuracies in the predictions in both main B-cell tumor entities
236 (mean sensitivities for training series = 0.97 and validation series =0.99) and B-cell
237 tumor subtypes (mean sensitivities for training series = 0.9 and validation series =
238 0.97). In some tumor subtypes, we obtained lower accuracy in the predictions mostly
239 due to small sample sizes. The script to easily implement the classifier is provided as
240 supplementary information (SI1). This epigenetic classifier may represent the basis for
241 a simple and accurate diagnostic tool for main B-cell tumors and B-cell tumor subtypes
242 with more complicated diagnosis such as subtypes of MCL or CLL.
243

244 **Patient-specific DNA methylation changes are associated with silent chromatin**
245 **without an impact on gene expression**

246 After having characterized entity-based sources of DNA methylation variability, we
247 next aimed at studying patient-specific changes within each tumor subtype (Fig. 1b,
248 level 4). To that end, we computed the total number and the number of hyper- and
249 hypomethylation changes in every single patient within each B-cell tumor subtype as
250 compared to HPC (Fig. 4a). As each B-cell tumor entity is derived from a distinct
251 cellular origin, this approach has the advantage of fixing a reference point for all B-cell
252 tumors, and the changes observed can be subsequently dissected into those
253 modulated in normal B-cell maturation and those taking place exclusively in the
254 context of neoplastic transformation (i.e. B-cell independent changes). Overall, we
255 found large differences in the numbers of DNA methylation changes per patient (Fig.
256 4a and Supplementary Table 8). To analyze whether some B-cell neoplasms show an
257 intrinsically more variable epigenome, we studied the degree of DNA methylation
258 variability in each group, but no differences were observed (Extended Data Fig. 4a).

259 The total number of altered CpGs per case was classified into four categories,
260 depending whether DNA methylation was gained or lost and whether it was
261 modulated or not during normal B cell development [6] (Fig. 2b and Fig. 4a). Regardless
262 of the cellular origin of each B-cell tumor entity, this analysis uncovered striking
263 correlations between the degree of B-cell related and B-cell independent DNA
264 methylation changes, a finding that was maintained when hypermethylation and
265 hypomethylation were studied separately (Fig. 4b and Extended Data Fig. 4b). This
266 association suggests that the overall DNA methylation burden of the tumor in each
267 individual patient may be shaped by a similar underlying phenomenon, which is
268 manifested in all four CpG categories. This statement is further supported by a
269 thorough annotation of the CpGs sites. Patient-specific CpGs that undergo
270 hypomethylation in the B-cell related and B-cell independent fractions are consistently
271 located in low CpG-content (open sea) low-signal heterochromatin, and the associated
272 genes are constitutively silent both in normal and neoplastic B cells (Fig. 4c-e and
273 Extended Data Fig. 4c-f). In the case of patient-specific hypermethylation, CpGs in both
274 fractions are located in promoter regions and CpG islands (CGIs) with H3K27me3-
275 repressed and poised-promoter chromatin states, and affect genes that remain silent
276 across normal differentiation and neoplastic transformation of B cells (Fig. 4f-h,
277 Extended Data Fig. 4c, g-i). These findings indicate that most DNA methylation
278 changes in B-cell tumor patients occur in silent chromatin regions in the absence of
279 concurrent phenotypic changes, suggesting that a mechanism independent from gene
280 regulation may underlie their overall DNA methylation landscape.
281 Beyond the classical role of DNA methylation as gene regulator, an accumulated body
282 of published evidence supports the concept that DNA methylation changes accumulate

283 during cell division (Fig. 4i) [33–39]. Studies in fibroblasts and hematopoietic stem cells
284 have reported that DNA methylation loss in late-replication heterochromatic regions
285 and DNA methylation gain of polycomb targets increase as cells proliferate without an
286 apparent impact on gene expression [33–35]. Furthermore, DNA methylation loss in
287 heterochromatic, late-replicating domains has been recently associated with mitotic
288 cell division in cancer [39]. Additionally, early studies detected preferential methylation
289 of CGI marked by H3K27me3 in cancer [40–42], and DNA methylation changes in regions
290 marked by H3K27me3 in embryonic stem cells have been recently related to mitotic
291 cell division in cancer [38].

292 In order to explore whether DNA methylation changes in silent regions reflect
293 the proliferative history of primary human B cells, we used DNA methylation data of an
294 *in vitro* differentiation model of primary NBCs into plasma cells (Fig. 4j) [43]. At days 4
295 and 6, different B cells were separated based on their proliferation history measured
296 by CFSE dilution. We evaluated the DNA methylation profile in repressed regions at
297 these time points and we detected the presence of hypermethylation at repressed
298 H3K27me3-containing regions and hypomethyation of low signal/H3K9me3-containing
299 heterochomatic regions in cells that have proliferated (Fig. 4k). This finding was
300 particularly marked at day 6, in which the gradual accumulation of DNA methylation
301 changes in silent regions was directly associated with their proliferative history (from
302 less divided P3 to highly divided P1). The genes associated to the measured CpGs did
303 not show any change of expression levels regardless of their methylation status, and
304 thus were unlikely to be related to the phenotype of the cells (Fig. 4l).

305 Collectively, all these data indicate that a great fraction of the patient-specific DNA
306 methylation changes in B-cell tumors accumulate at H3K27me3 and low

307 signal/H3K9me3 regions during cell division without affecting gene expression levels, a
308 finding that is consistent with an epigenetic mitotic clock.
309

310 **Development and validation of an epigenetic mitotic clock reflecting the proliferative**
311 **history of normal and neoplastic B cells**

312 We performed a step-wise selection of CpGs whose methylation status would reflect
313 the cell mitotic activity regardless of its normal or neoplastic nature (Methods). Using
314 this strategy, we finally retained 184 CpGs in stable polycomb regions that tend to gain
315 DNA methylation and 1,164 CpGs in constitutive heterochromatin that tend to lose
316 DNA methylation during normal B-cell maturation and neoplastic transformation (Fig.
317 5a, Supplementary Table 9 and SI2). We next constructed two scores, one for
318 hypermethylation and one for hypomethylation, which we respectively named
319 epiCMIT-hyper and epiCMIT-hypo, ranging from 0 to 1 depending on low or high
320 proliferative histories, respectively (Methods). Remarkably, epiCMIT-hypo contained a
321 significant proportion of solo-WCGW CpGs, which have been recently associated with
322 human mitotic cell division (N=1e5 permutations, pval<0.0001) [39]. As normal B cell
323 differentiation entails cell division, we initially evaluated both scores in normal B cells
324 and observed an expected but strikingly high correlation (R=0.96, pval<0.001), with B-
325 cell subpopulations distributed according to their maturation state (and thus according
326 to their accumulated proliferative history during B-cell differentiation) (Fig. 5b, left
327 panel). This finding suggests that mitotic cell division in normal B cells leaves both
328 hyper- and hypomethylated imprints. This high correlation between the two scores
329 was also observed for MCL, CLL and DLBCL (Fig. 5b) but not for ALL and MM. The
330 epiCMIT-hyper was greater than the epiCMIT-hypo in ALL samples, and the opposite

331 scenario was observed in MM. As these two neoplasias originate from B cells at the
332 two extremes of the maturation spectrum, our data suggests that the impact of cell
333 division onto the DNA methylome may be different depending on the maturation state
334 of the cellular origin. Upon neoplastic transformation, dividing precursor B cells do not
335 seem to acquire broad hypomethylation in heterochromatin but rather
336 hypermethylation in polycomb-repressed regions. In contrast, neoplastic plasma cells
337 acquire widespread hypomethylation in heterochromatin and virtually lack
338 hypermethylation of polycomb-repressed regions. As previously shown in Extended
339 Data Fig. 2d, this phenomenon may be related to the differential expression of DNMTs
340 of the respective cells of origin of ALL and MM. We then took into consideration both
341 scores to derive a unique epiCMIT score (Fig. 5c). The epiCMIT reflects the relative
342 accumulation of mitotic cell divisions of a particular sample and is able to capture
343 different tendencies in gaining or losing DNA methylation during mitotic cell division,
344 as it respectively happens in ALL and MM patients (Fig. 5b). Finally, epiCMIT cannot be
345 affected by different distribution of cell cycle phases in tumor samples, since the DNA
346 methylome remains rather stable during the whole cell cycle [44].

347 The applicability of the epiCMIT as mitotic clock was validated through several
348 perspectives. First, we compared it with two previously reported hypermethylation-
349 based mitotic clocks called epiTOC and MiAge [38,45] (Supplementary Table 8). The
350 epiCMIT showed excellent correlations with these clocks in B-cell neoplasms that tend
351 to acquire polycomb-related hypermethylation (e.g. mostly ALL, but also DLBCL and
352 MCL); a moderate correlation in the case of CLL, which acquires more hypo- than
353 hypermethylation (Fig. 4a and Extended Data Fig. 4b), and a total lack of correlation in
354 the case of MM, which mostly loses DNA methylation (Fig. 5d and Extended Data Fig.

355 5a). Identical observations were obtained comparing the epiCMIT and the widely-
356 reported CpG island methylator phenotype (CIMP) in human cancer [46] calculated as
357 previously proposed [47] (Supplementary Table 8). Therefore, our analysis also reveals
358 that the classical CIMP score in reality may represent a measure of the mitotic history
359 of the cells. A potentially confusing aspect related to epiCMIT is the fact that DNA
360 methylation has mainly been used to measure the chronological age of an individual
361 [48,49]. In order to study the relationship between mitotic and aging DNA methylation
362 clocks, we first compared the epiCMIT in naïve B cells from infant, adults and elderly
363 donors and showed that epiCMIT was stable in all age ranges regardless of donor's age
364 (Extended Data Fig. 5b). We next applied the most popular chronological age clock
365 proposed by Horvath [50] and showed that the age of donors was predicted with high
366 accuracy. This result indicates that, although both Horvath and epiCMIT scores are
367 based on DNA methylation, they measure distinct biological phenomena (Extended
368 Data Fig. 5b). Furthermore, the epiCMIT was highly variable in pediatric ALL samples
369 with a minimum age range (Extended Data Fig. 5c). Collectively, these analyses
370 indicate that the epiCMIT is a more universal mitotic clock than hypermethylation-
371 based clocks since it captures the proliferative history of neoplasms regardless of their
372 tendency of gaining or losing DNA methylation during mitotic cell division.

373 Second, the epiCMIT was validated using the *in vitro* B-cell differentiation
374 model shown in Fig. 4j, in which we observed that epiCMIT increases with the
375 proliferative history of the cells without altering gene expression levels of the epiCMIT-
376 associated genes (Fig. 5e). Third, using WGS data from 138 CLL patients from our
377 cohort [10,17,51,52], we observed that the epiCMIT was highly correlated with the total
378 number of somatic mutations (Extended Data Fig. 5d). We next extracted mutational

379 signatures as recently described [53] (Extended Data Fig. 5e) and observed significant
380 correlations with mutational signatures SBS1 and SBS5 (Fig. 5f), that have been
381 recently described as mitotic-like mutational processes [54]. We also identified a
382 significant link between the epiCMIT and the non-canonical AID signature (SBS9) (Fig.
383 5f) [52,55] in IGHV mutated CLL, possibly reflecting rounds of cell divisions in germinal
384 center B cells before differentiation into memory B cells and malignant transformation.
385 Fourth, although the epiCMIT reflects the proliferative history of the cell rather than
386 the actual proliferative status of the samples (e.g, bmPC have high epiCMIT and do not
387 proliferate, Fig. 5c), a relationship between epiCMIT and cell proliferation is expected
388 (more proliferative history implies overall more proliferation, although also depends
389 on time). Accordingly, MCL cases showing higher Ki-67 (a proliferation marker) also
390 had higher epiCMIT than cases with moderate Ki-67 expression (Extended Data Fig 5f).
391 Furthermore, GSEA analyses in ALL and CLL cases with high and low epiCMIT revealed
392 that cases with high epiCMIT showed higher expression of genes related with cell
393 proliferation (Extended Data Fig. 5g, h). Thus, these data suggest that cases with higher
394 proliferative history also seem to have higher proliferation at the time of sampling.
395 Collectively, the four lines of evidence presented above support that the epiCMIT may
396 represent a *bona fide* measure of the relative number of cumulative mitotic cell
397 divisions that normal and neoplastic B cells have undergone since the uncommitted
398 hematopoietic cell stage.
399
400
401

402 **The epiCMIT is a strong independent variable predicting clinical behavior in B-cell**
403 **tumors**

404 In normal B-cell maturation, the epiCMIT gradually augments as B cells proliferate
405 during cell differentiation, an increase that is particularly marked in proliferative GC B
406 cells (Fig. 5c). In neoplastic B cells, however, the interpretation of the epiCMIT is less
407 trivial and must be divided into two components: the epiCMIT of the cell of origin and
408 the epiCMIT acquired in the course of the neoplastic transformation and progression
409 (Extended Data Fig. 6a). Therefore, the relative epiCMIT must be compared among
410 patients from entities arising from the same B-cell maturation stage and should be a
411 dynamic variable during cancer progression. Thus, we compared the epiCMIT in two
412 paradigmatic transitions between precursor conditions and overt cancer, i.e.
413 monoclonal gammopathy of undetermined significance (MGUS) and MM, as well as
414 monoclonal B cell lymphocytosis (MBL) and CLL categorized according to their cellular
415 origin. This analysis showed an overall lower epiCMIT in precursor lesions compared
416 with overt cancer (Fig. 6a, upper panels), as would be expected due to their increased
417 leukemia-specific proliferative history. In line with this finding, the epiCMIT increased
418 in paired CLLs at diagnosis and progression before treatment as well as in sequential
419 ALL samples at diagnosis, first relapse and second relapse (Fig. 6a, lower panels). These
420 results suggest that the epiCMIT evolves together with clinical progression.

421 Based on our previous observations, we next wondered whether the epiCMIT
422 could be useful to predict the clinical behavior of B-cell neoplasms. We analyzed the
423 relationship of the epiCMIT with clinical outcome in specific B-cell tumor subtypes
424 based on their shared cytogenetics (i.e. ALL) or cell of origin (i.e. MCL, CLL and DLBCL),
425 and thus having a similar ground state B-cell specific proliferative history (Extended

426 Data Fig. 6a). In ALL, high epiCMIT was consistently associated with longer relapse-free
427 and overall survival (Fig. 6b and Extended Data Fig. 6b) of the patients within each
428 cytogenetic subtype, and thus better clinical outcome. Next, we additionally showed
429 epiCMIT as significant variable in a multivariate regression Cox model with epiCMIT as
430 continuous variable together with cytogenetic subtypes (Fig. 6b). This observation is in
431 line with published evidence showing that a higher CIMP, which is highly correlated
432 with the epiCMIT in ALL (Fig. 5d), is a good prognosis factor in this disease [56,57]. In
433 contrast to ALL, the opposite clinical scenario was observed in mature B-cell
434 neoplasms. In each of the CLL subtypes, a high epiCMIT was strongly associated with a
435 worse prognosis using time to treatment (TTT) as end-point variable, both from
436 sampling time (Fig. 6c, left panel) and in cases whose sample was obtained close to
437 diagnosis (Extended Data Fig. 6c, left panel). Additionally, a multivariate Cox regression
438 model for TTT revealed that epiCMIT as continuous variable was a highly significant
439 variable conferring dismal prognosis together with age, number of driver genetic
440 alterations [17,58] and epigenetic subgroups [10,11,59] (Fig. 6c right panel and Extened Data
441 Fig. 6c). The epiCMIT maintained its independent prognostic value using overall
442 survival as end-point variable, although its effect was moderate (Extended Data Fig.
443 6d). These findings were widely confirmed in an additional series of 136 CLLs treated
444 with chemo-immunotherapy (Fig. 6d and Extended Data Fig. 6c, d right panels). In the
445 case of MCL, the epiCMIT showed an independent poor prognostic impact in the two
446 cell-of-origin subtypes (C1 and C2), an observation that was confirmed in an extended
447 series in the more aggressive and prevalent C1 group (Fig. 6e, f). Lastly, although the
448 sample size was limited and requires further studies, our data suggest that high
449 epiCMIT could also represent a poor prognostic variable within the two cell-of-origin

450 DLBCL subtypes (Extended Data Fig. 6e). Overall, these results suggest that an
451 increased proliferative history of the neoplastic clone in precursor B-cell neoplasms at
452 diagnosis predicts for a better disease-free survival, as previously published in studies
453 analyzing the CIMP score [56,57]. In sharp contrast, in mature B-cell neoplasms, which are
454 overall less proliferative than ALL, neoplastic clones with high proliferation history
455 seem to predict for future proliferative capacity and consistently show worse clinical
456 outcomes.
457

458 **epiCMIT is associated with specific genetic driver alterations in CLL**

459 We next sought to assess which CLL driver alterations may confer a proliferative
460 advantage to neoplastic cells, and subsequently a higher epiCMIT. To that end, we
461 exploited 477 CLL samples in which we had DNA methylation data and whole exome
462 sequencing (WES) (Fig. 7a). We initially depicted all driver genetic changes in each CLL
463 subtype divided in high and low epiCMIT (Extended Data Fig. 7a). Next, we
464 interrogated the levels of epiCMIT in patients with each driver genetic alteration (with
465 at least 4 mutated patients). We performed the analyses in all CLL patients (showed as
466 Global in Fig. 7b) and then within each epigenetic subgroup separately (Fig. 7b,
467 Extended Data Fig. 7b and Methods). We showed significant and positive associations
468 of epiCMIT with 24 genetic driver alterations affecting the main signaling pathways
469 altered in CLL (Fig. 7b,c) [17,58]. The majority of these genetic alterations have been
470 previously linked to an adverse clinical behavior of patients, such as *NOTCH1, TP53,*
471 *SF3B1, ATM, BIRC3 or EGR2*. Interestingly, epiCMIT showed a clear association with a
472 new and recently identified non-coding genetic driver in CLL, the U1 spliceosomal RNA,
473 a finding that may help in explaining its suggested poor prognostic impact [60].

474 Remarkably, the presence of some genetic alterations was associated with high
475 epiCMIT indistinctly in all patients, such as *TP53*, while others were particularly
476 associated with epiCMIT within CLL subgroups, such as *NOTCH1* in n-CLL and i-CLL,
477 *SF3B1* in i-CLL, and del(13q) and tri12 in m-CLL.
478 Collectively, these results suggest that the well-established clinical impact of genetic
479 alterations in CLL may be explained by their association with a high proliferative
480 potential, being this association different for certain genetic alterations depending on
481 the maturation state of the cellular origin.
482

483 **Discussion**

484 In this study, we have followed a systematic approach to dissect the sources of DNA
485 methylation variability of B-cell neoplasms in the context of the normal B-cell
486 differentiation program. Overall, we found that the methylation levels of 88% of all
487 CpGs are modulated in normal and/or neoplastic B cells, suggesting that the human
488 DNA methylome is even more dynamic than previously appreciated [6,61,62]. The
489 extensive DNA methylation variability among different B-cell neoplasms is in part
490 related to imprints of normal B-cell development. This phenomenon has been
491 previously observed and has led to a more accurate classification not only of B-cell
492 neoplasms [9–11,59], but also of solid tumors [2,63,64]. In addition to this epigenetic link to
493 normal cell maturation, each B-cell neoplasm also shows disease-specific hyper- and
494 hypomethylation. Of particular interest are the disease-specific *de novo*
495 hypomethylation signatures in active regulatory regions, which were associated with
496 binding of TFs involved in the pathogenesis of each respective B-cell tumor. This
497 phenomenon was particularly marked in ALL, MCL and CLL, whose *de novo*

498 hypomethylation was enriched in regulatory elements [23–25]. Unexpectedly, although
499 DLBCL and MM pathogenesis has been linked to TFs, *de novo* hypomethylation was
500 depleted in regulatory elements containing TF binding sites. In these two malignancies,
501 we detected few binding sites of TFs with potential involvement in the diseases.
502 However, we did not detect classical TFs such as BCL6 in DLBCL or IRF4 in MM, possibly
503 because they are key players during B-cell differentiation and their binding sites may
504 already be hypomethylated in the normal B-cell counterparts of DLBCL and MM.

505 In spite of the importance of DNA methylation in regulatory regions, we
506 identified that the majority of DNA methylation changes in B-cell neoplasms, and
507 especially those related to inter-patient variability within each B-cell tumor subtype,
508 are located in inactive chromatin. The magnitude of this inter-patient variability
509 affecting normal B-cell related and -independent regions is highly correlated,
510 suggesting that the cause of these DNA methylation changes is a phenomenon that
511 takes place both in normal and neoplastic B cells. Compelling published evidence as
512 well as new data presented in our study support the notion that mitotic cell division
513 leaves transcriptionally-inert epigenetic imprints onto the DNA. These take place in the
514 form of hypomethylation of heterochromatin and hypermethylation of polycomb-
515 repressed regions. This knowledge has recently led to the concept of using DNA
516 methylation as a mitotic clock [38,45], and has been recently used by single cell DNA
517 methylation data to track the evolutionary trajectory of CLL [65]. Notably, methylation-
518 based mitotic clocks seem to capture different biological information from aging clocks
519 (Extended Data Fig. 5b) [48]. As far as we are aware, there are two published mitotic
520 clocks [38,45] and both are hypermethylation-based. Furthermore, our analyses suggest
521 that the classical CIMP phenomenon in cancer [46] may indeed represent another

522 hypermethylation-based mitotic clock. However, only using hypermethylation to
523 determine the mitotic clock is insufficient to capture the mitotic activity of the cells, as
524 some neoplasms may not acquire hypermethylation upon cell division but rather
525 hypomethylation. For instance, neoplastic plasma cells in MM do not tend to acquire
526 polycomb-related hypermethylation. Thus, using hypermethylation to determine the
527 mitotic history of MM cells would incongruently lead to the conclusion that they have
528 not proliferated beyond their cellular origin (Fig. 5d and Extended Data Fig. 5a).
529 Therefore, to circumvent possible misleading interpretations, we generated a more
530 broadly applicable mitotic clock (called epiCMIT) that takes into consideration both
531 hyper- and hypomethylation. Importantly, epiCMIT aims at capturing the entire mitotic
532 history of cells, including cell division associated both with normal development as well
533 as neoplastic transformation and progression (Extended Data Fig. 6a). Thus, the
534 epiCMIT must not be compared among tumors arising from different normal
535 counterparts but its relative magnitude must be studied in those arising from a
536 particular maturation stage. Within each of these subgroups, the relative epiCMIT has
537 a profound independent prognostic value from other well-established clinical
538 variables. Increased epiCMIT is associated with worse clinical outcome within CLL and
539 MCL subgroups, thus indicating that superior proliferative history before treatment
540 seems to determine future proliferative capacity and is thus associated with worse
541 clinical outcome. Strikingly, we consistently found the opposite pattern in precursor
542 ALL subgroups, a finding in line with recent reports showing that the presence of CIMP
543 (a hypermethylation-based mitotic clock) is associated with better clinical outcome
544 [56,57]. These results suggest that children having more proliferative ALL cells at diagnosis

545 (and thus a larger proliferative history) are more prone to be cured with high intensive
546 chemotherapy regimens, [66], which cannot be administrated in elderly patients.

547 Finally, in order to identify the genetic lesions associated with higher
548 proliferative history in CLL, we exploited our thorough genetic characterization of our
549 cases to study the relationship between driver genetic alterations and epiCMIT. We
550 identified 24 driver genetic alterations that may confer a higher proliferative capacity
551 and thus are associated with higher epiCMIT or methylation evolution [67]. These genetic
552 alterations were distributed throughout the main altered signaling pathways in CLL
553 and were manifested differently in distinct CLL subgroups based on their cellular origin
554 (Fig. 7b, c). This finding suggests that specific mutations may predispose to a higher
555 proliferative advantage depending on the maturation stage and (epi)genetic makeup
556 of the cellular origin.

557 In summary, our comprehensive epigenetic evaluation of normal and neoplastic
558 B cells at different maturation stages uncovers multiple new insights into the biological
559 roles of DNA methylation in cancer, an analytic approach that may also benefit our
560 understanding of other cancers. From a clinical perspective, DNA methylation may
561 provide a holistic diagnostic and prognostic approach to B-cell neoplasms. Particularly,
562 we defined an accurate and easy-to-implement pan-B-cell tumor diagnostic tool and
563 generated a mitotic clock reflecting the proliferative history of the neoplastic cells of
564 each patient to estimate their clinical risk, which shall represent a valuable asset in the
565 precision medicine era.

566

ACKNOWLEDGEMENTS

This research was funded by the European Union's Seventh Framework Programme through the Blueprint Consortium (grant agreement 282510), Generalitat de Catalunya Suport Grups de Recerca AGAUR 2017-SGR-1142 (to E.C.) and 2017-SGR-736 (to J.I.M.-S.), Ministerio de Ciencia, Innovación y Universidades of the Spanish Government (MCIU), Grants RTI2018-094274-B-I00 (to E.C.) and SAF2017-86126-R (to J.I.M.-S.) as well as Proyecto Medicina Personalizada PERMED (Grant PMP15/00007), which is part of Plan Nacional de I+D+I and is co-financed by the ISCIII-Sub-Directorate General for Evaluation and the European Regional Development Fund (FEDER-"Una manera de Hacer Europa"), CIBERONC (CB16/12/00225, CB16/12/00334, CB16/12/00236, and CB16/12/00489), the Accelerator award CRUK/AIRC/AECC joint funder-partnership, research funding from Fondo de Investigaciones Sanitarias, Instituto de Salud Carlos III PI17/01061 (SB), Ministerio de Ciencia, Innovación y Universidades (MCIU), RTI2018-094274-B-I00, SAF2015-64885-R (EC), the NIH grant number 1 P01CA229100 (EC), and the European Regional Development Fund "Una manera de fer Europa", CERCA Programme/Generalitat de Catalunya. FN is supported by a pre-doctoral fellowship of the Ministerio de Economía y Competitividad (MINECO, BES-2016-076372). E.C. is an Academia Researcher of the "Institució Catalana de Recerca i Estudis Avançats" (ICREA) of the Generalitat de Catalunya. This work was partially developed at the Centro Esther Koplowitz (CEK, Barcelona, Spain). We thank Francesco Maura for his help with the analysis of mutational signatures.

591 **AUTHOR CONTRIBUTIONS**

592 Investigator contributions were as follows: T.B., J.N., Y.N-Z., G.L., A.R-D., S.M., R.O.,
593 G.C., M.K., A.C-Q., L.S-T., J.W., J.L., E.G., S.B., P.J., X.A., F.P., C.L-O., X.S.P., C.C.O., T.Z.,
594 J.D., A.L-G. and E.C. contributed to sample biological and/or clinical annotation; M.D-
595 F., G.C., F.N. and R.B. performed methylome, ChIP-Seq, transcriptome, genetic and/or
596 statistical analyses; R.R., M.P. and D.T. provided computational support. M.D-F. and
597 J.I.M.-S. participated in the study design. M.D-F., F.N., R.B., T.B., J.D., A.L-G., E.C. and
598 J.I.M.-S. participated in data interpretation. J.I.M.-S. directed the research and wrote
599 the manuscript together with M.D-F.

600

601 **COMPETING INTERESTS**

602 The authors declare no competing interests.

603

604

605

606

607
608 **References**

609 1. Roy, N. & Hebrok, M. Regulation of Cellular Identity in Cancer. *Dev. Cell* **35**, 674–
610 84 (2015).

611 2. Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular
612 Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304.e6
613 (2018).

614 3. Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, T. J. *WHO*
615 *Classification of Tumours of Haematopoietic and Lymphoid Tissues*.
616 (International Agency for Research on Cancer (IARC), 2017).

617 4. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21
618 (2002).

619 5. Luo, C., Hajkova, P. & Ecker, J. R. Dynamic DNA methylation: In the right place at
620 the right time. *Science* **361**, 1336–1340 (2018).

621 6. Kulis, M. *et al.* Whole-genome fingerprint of the DNA methylome during human
622 B cell differentiation. *Nat. Genet.* **47**, 746–756 (2015).

623 7. Nordlund, J. *et al. Genome-wide signatures of differential DNA methylation in*
624 *pediatric acute lymphoblastic leukemia. Genome biology* **14**, (2013).

625 8. Lee, S.-T. *et al.* Epigenetic remodeling in B-cell acute lymphoblastic leukemia
626 occurs in two tracks and employs embryonic stem cell-like signatures. *Nucleic*
627 *Acids Res.* **43**, 2590–602 (2015).

628 9. Queirós, A. C. *et al.* Decoding the DNA Methylome of Mantle Cell Lymphoma in
629 the Light of the Entire B Cell Lineage. *Cancer Cell* **30**, 806–821 (2016).

630 10. Kulis, M. *et al.* Epigenomic analysis detects widespread gene-body DNA
631 hypomethylation in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 1236–1242
632 (2012).

633 11. Oakes, C. C. *et al.* DNA methylation dynamics during B cell maturation underlie a
634 continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat. Genet.*
635 (2016). doi:10.1038/ng.3488

636 12. Kretzmer, H. *et al.* DNA methylome analysis in Burkitt and follicular lymphomas
637 identifies differentially methylated regions linked to somatic mutation and
638 transcriptional control. *Nat. Genet.* (2015). doi:10.1038/ng.3413

639 13. Shaknovich, R. *et al.* DNA methylation signatures define molecular subtypes of
640 diffuse large B-cell lymphoma. *Blood* **116**, e81-9 (2010).

641 14. Agirre, X. *et al.* Whole-epigenome analysis in multiple myeloma reveals DNA
642 hypermethylation of B cell-specific enhancers. *Genome Res.* **25**, 478–87 (2015).

643 15. Kaiser, M. F. *et al.* Global methylation analysis identifies prognostically
644 important epigenetically inactivated tumor suppressor genes in multiple
645 myeloma. *Blood* **122**, 219–226 (2013).

646 16. Oakes, C. C. & Martin-Subero, J. I. Insight into origins, mechanisms & utility

647 of DNA methylation in B cell malignancies. *Blood* **132**, blood-2018-02-692970
648 (2018).

649 17. Puente, X. S. *et al.* Non-coding recurrent mutations in chronic lymphocytic
650 leukaemia. *Nature* (2015). doi:10.1038/nature14666

651 18. Karube, K. *et al.* Integrating genomic alterations in diffuse large B-cell lymphoma
652 identifies new relevant pathways and potential therapeutic targets. *Leukemia*
653 675–684 (2017). doi:10.1038/leu.2017.251

654 19. Stunnenberg, H. G., International Human Epigenome Consortium & Hirst, M.
655 The International Human Epigenome Consortium: A Blueprint for Scientific
656 Collaboration and Discovery. *Cell* **167**, 1145–1149 (2016).

657 20. Viré, E. *et al.* The Polycomb group protein EZH2 directly controls DNA
658 methylation. *Nature* **439**, 871–874 (2006).

659 21. Stadler, M. B. *et al.* Corrigendum: DNA-binding factors shape the mouse
660 methylome at distal regulatory regions. *Nature* **484**, 550–550 (2012).

661 22. Burger, L., Gaidatzis, D., Schübeler, D. & Stadler, M. B. Identification of active
662 regulatory regions from DNA methylation data. *Nucleic Acids Res.* **41**, (2013).

663 23. Somasundaram, R., Prasad, M. A. J., Ungerbäck, J. & Sigvardsson, M.
664 Transcription factor networks in B-cell differentiation link development to acute
665 lymphoid leukemia. *Blood* **126**, 144–152 (2015).

666 24. Sánchez-Tilló, E. *et al.* The EMT activator ZEB1 promotes tumor growth and
667 determines differential response to chemotherapy in mantle cell lymphoma. *Cell*
668 *Death Differ.* **21**, 247–257 (2014).

669 25. Wolf, C. *et al.* NFATC1 activation by DNA hypomethylation in chronic
670 lymphocytic leukemia correlates with clinical staging and can be inhibited by
671 ibrutinib. *Int. J. Cancer* **142**, 322–333 (2018).

672 26. Blonska, M. *et al.* Jun-regulated genes promote interaction of diffuse large B-cell
673 lymphoma with the microenvironment. *Blood* **125**, 981–991 (2015).

674 27. Huerta-Yepez, S. *et al.* Overexpression of Yin Yang 1 in bone marrow-derived
675 human multiple myeloma and its clinical significance. *Int. J. Oncol.* **45**, 1184–
676 1192 (2014).

677 28. Sprynski, A. C. *et al.* Insulin is a potent myeloma cell growth factor through
678 insulin/IGF-1 hybrid receptor activation. *Leukemia* **24**, 1940–1950 (2010).

679 29. Riz, I. & Hawley, R. G. Increased expression of the tight junction protein
680 TJP1/ZO-1 is associated with upregulation of TAZ-TEAD activity and an adult
681 tissue stem cell signature in carfilzomib-resistant multiple myeloma cells and
682 high-risk multiple myeloma patients. *Oncoscience* **4**, 79–94 (2017).

683 30. Herath, N. I., Rocques, N., Garancher, A., Eychène, A. & Pouponnot, C. GSK3-
684 mediated MAF phosphorylation in multiple myeloma as a potential therapeutic
685 target. *Blood Cancer J.* **4**, e175–e175 (2014).

686 31. Navarro, A. *et al.* Improved classification of leukemic B-cell lymphoproliferative
687 disorders using a transcriptional and genetic classifier. *Haematologica* **102**, 360–

688 363 (2017).

689 32. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by
690 gene expression profiling. *Nature* **403**, 503–511 (2000).

691 33. Aran, D., Toperoff, G., Rosenberg, M. & Hellman, A. Replication timing-related
692 and gene body-specific methylation of active human genes. *Hum. Mol. Genet.*
693 **20**, 670–680 (2011).

694 34. Beerman, I. *et al.* Proliferation-dependent alterations of the DNA methylation
695 landscape underlie hematopoietic stem cell aging. *Cell Stem Cell* **12**, 413–25
696 (2013).

697 35. Landan, G. *et al.* Epigenetic polymorphism and the stochastic formation of
698 differentially methylated regions in normal and cancerous tissues. *Nat. Genet.*
699 **44**, 1207–14 (2012).

700 36. Siegmund, K. D., Marjoram, P., Woo, Y.-J., Tavaré, S. & Shibata, D. Inferring
701 clonal expansion and cancer stem cell dynamics from DNA methylation patterns
702 in colorectal cancers. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 4828–4833 (2009).

703 37. Spencer, D. H. *et al.* CpG Island Hypermethylation Mediated by DNMT3A Is a
704 Consequence of AML Progression. *Cell* **168**, 801-816.e13 (2017).

705 38. Yang, Z. *et al.* Correlation of an epigenetic mitotic clock with cancer risk.
706 *Genome Biol.* **17**, 205 (2016).

707 39. Zhou, W. *et al.* DNA methylation loss in late-replicating domains is linked to
708 mitotic cell division. *Nat. Genet.* **50**, 591–602 (2018).

709 40. Schlesinger, Y. *et al.* Polycomb-mediated methylation on Lys27 of histone H3
710 pre-marks genes for de novo methylation in cancer. *Nat. Genet.* **39**, 232–236
711 (2007).

712 41. Ohm, J. E. *et al.* A stem cell–like chromatin pattern may predispose tumor
713 suppressor genes to DNA hypermethylation and heritable silencing. *Nat. Genet.*
714 **39**, 237–242 (2007).

715 42. Widschwendter, M. *et al.* Epigenetic stem cell signature in cancer. *Nat. Genet.*
716 **39**, 157–158 (2007).

717 43. Caron, G. *et al.* Cell-Cycle-Dependent Reconfiguration of the DNA Methylome
718 during Terminal Differentiation of Human B Cells into Plasma Cells. *Cell Rep.* **13**,
719 1059–71 (2015).

720 44. Vandiver, A. R., Idrizi, A., Rizzardi, L., Feinberg, A. P. & Hansen, K. D. DNA
721 methylation is stable during replication and cell cycle arrest. *Sci. Rep.* **5**, 1–8
722 (2015).

723 45. Youn, A. & Wang, S. The MiAge Calculator: a DNA methylation-based mitotic age
724 calculator of human tissue types. *Epigenetics* **13**, 192–206 (2018).

725 46. Issa, J. CpG island methylator phenotype in cancer. *Nat. Rev. Cancer* **4**, 988–93
726 (2004).

727 47. Sánchez-Vega, F., Gotea, V., Margolin, G. & Elnitski, L. Pan-cancer stratification
728 of solid human epithelial tumors and cancer cell lines reveals commonalities and

729 tissue-specific features of the CpG island methylator phenotype. *Epigenetics and*
730 *Chromatin* **8**, 1–24 (2015).

731 48. Field, A. E. *et al.* DNA Methylation Clocks in Aging: Categories, Causes, and
732 Consequences. *Mol. Cell* **71**, 882–895 (2018).

733 49. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic
734 clock theory of ageing. *Nat. Rev. Genet.* (2018). doi:10.1038/s41576-018-0004-3

735 50. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol*
736 **14**, R115 (2013).

737 51. Ferreira, P. G. *et al.* Transcriptome characterization by RNA sequencing
738 identifies a major molecular and clinical subdivision in chronic lymphocytic
739 leukemia Gene Regulation Stem Cells and Cancer Programme, Centre for
740 Genomic Regulation (CRG). 212–226 (2014). doi:10.1101/gr.152132.112

741 52. Puente, X. S. *et al.* Whole-genome sequencing identifies recurrent mutations in
742 chronic lymphocytic leukaemia. *Nature* **475**, 101–105 (2011).

743 53. Maura, F. et al. A practical guide for mutational signature analysis in
744 hematological malignancies. *Nat. Commun.* (2019). doi:10.1038/s41467-019-
745 11037-8

746 54. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells.
747 *Nat. Genet.* **47**, 1402–7 (2015).

748 55. Alexandrov, L. *et al.* The Repertoire of Mutational Signatures in Human Cancer.
749 *bioRxiv* (2018). doi:https://doi.org/10.1016/j.cell.2011.02.013

750 56. Borssén, M. *et al.* DNA methylation holds prognostic information in relapsed
751 precursor B-cell acute lymphoblastic leukemia. *Clin. Epigenetics* **10**, 31 (2018).

752 57. Sandoval, J. *et al.* Genome-wide DNA methylation profiling predicts relapse in
753 childhood B-cell acute lymphoblastic leukaemia. *Br. J. Haematol.* **160**, 406–9
754 (2013).

755 58. Landau, D. A. *et al.* Mutations driving CLL and their evolution in progression and
756 relapse. *Nature* **526**, 525–30 (2015).

757 59. Queirós, a C. *et al.* A B-cell epigenetic signature defines three biological
758 subgroups of chronic lymphocytic leukemia with clinical impact. *Leukemia* 598–
759 605 (2015). doi:10.1038/leu.2014.252

760 60. Shuai, S. *et al.* The U1 spliceosomal RNA is recurrently mutated in multiple
761 cancers. *Nature* **574**, 712–716 (2019).

762 61. Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human
763 genome. *Nature* **500**, 477–81 (2013).

764 62. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread
765 epigenomic differences. *Nature* **462**, 315–322 (2009).

766 63. Rodríguez-Paredes, M. *et al.* Methylation profiling identifies two subclasses of
767 squamous cell carcinoma related to distinct cells of origin. *Nat. Commun.* **9**,
768 (2018).

769 64. Bormann, F. *et al.* Cell-of-Origin DNA Methylation Signatures Are Maintained
770 during Colorectal Carcinogenesis. *Cell Rep.* **23**, 3407–3418 (2018).

771 65. Gaiti, F. *et al.* Epigenetic evolution and lineage histories of chronic lymphocytic
772 leukaemia. *Nature* (2019). doi:10.1038/s41586-019-1198-z

773 66. Rhein, P. *et al.* Gene expression shift towards normal B cells, decreased
774 proliferative capacity and distinct surface receptors characterize leukemic blasts
775 persisting during induction therapy in childhood acute lymphoblastic leukemia.
776 *Leukemia* **21**, 897–905 (2007).

777 67. Oakes, C. C. *et al.* Evolution of DNA Methylation Is Linked to Genetic Aberrations
778 in Chronic Lymphocytic Leukemia. *Cancer Discov.* **4**, 348–361 (2014).

779

780

781
782 **Figure legends**

783 **Fig. 1: Shared DNA methylation dynamics in B-cell tumors.**

784 **a,** Number of normal and neoplastic B-cells included in the study with tumor cell
785 content of at least 60% (additional samples for validations are used in Fig. 3 and Fig. 6,
786 which are later detailed). HPC, hematopoietic precursor cells; pre-B, precursor B-cell
787 and immature B cells; NBC, naïve B cells; GC, germinal center B cells; MBC, memory B
788 cells; tPC, tonsillar plasma cells; bmPC, bone-marrow plasma cells; ALL, acute
789 lymphoblastic leukemia; MCL, mantle cell lymphoma; CLL, chronic lymphocytic
790 leukemia; DLBCL, Diffuse large B cell lymphoma; MM, multiple myeloma; BM, bone
791 marrow; PB, peripheral blood; LN, lymph node.

792 **b**, Different levels of DNA methylation variability addressed in the study.

793 **c,** Heatmaps showing shared DNA hyper- (top) and hypomethylation (bootm) in B-cell
794 tumors.

795 **d,** Chromatin state enrichments of regions sharing DNA hyper- and hypomethylation
796 (all CpGs of 450k array were used as background). ActProm, Active promoter;
797 WkProm, Weak promoter; StrEnh1, Strong enhancer 1 (promoter-related); StrEnh2,
798 Strong enhancer 2; WkEnh, Weak enhancer; TxnTrans, Transcription transition;
799 TxnElong, Transcription elongation; WkTxn, Weak transcription; PoisProm, Poised
800 promoter; H3K27me3, Polycomb-repressed region; H3K9me3, H3K9me3
801 heterochromatin; Het;LowSign, Het;LowSign heterochtomatin.

802 **e,** Overlap of the target genes of stably methylated and unmethylated CpGs.

803 **f,** Gene expression percentile within each sample of genes showing stable hyper- and
804 hypomethylation.

805

806 **Fig. 2: Disease-specific hypomethylation at H3K27ac regions is associated with**
807 **specific transcription factor bindings and differential gene expression.**

808 **a,** Principal component analysis of normal and neoplastic B-cells, including ALL, MCL,
809 CLL, DLBCL and MM. ALL, acute lymphoblastic leukemia; MCL, mantle cell lymphoma;
810 CLL, chronic lymphocytic leukemia; DLBCL, Diffuse large B cell lymphoma; MM,
811 multiple myeloma.

812 **b,** Number of *de novo* DNA methylation changes in each B-cell tumor entity.

813 **c,** Heatmap showing entity-specific hypomethylation and the number of CpGs falling at
814 regulatory regions. Per each B-cell tumor, the same number of *de novo* CpGs was
815 randomly chosen from the 450K array and interrogated the percentage falling at
816 regulatory regions.

817 **d,** Transcription factor binding site predictions for *de novo* hypomethylated CpGs
818 falling at regulatory regions in c, for each B-cell tumor.

819 **e,** Differential gene expression percentiles among B-cell tumors for genes showing
820 specific-hypomethylation at regulatory regions.

821

822 **Fig. 3: Accurate classification of 14 clinico-biological subtypes of B cell neoplasms**
823 **using epigenetic biomarkers.**

824 **a,** Heatmaps for the CpGs used for the pan B-cell cancer classifier. The classifier
825 consists of two steps: in the first step (1) an unknown B-cell tumor can be predicted
826 into ALL, MCL, CLL, DLBCL or MM, and subsequently (second step, using one of the
827 predictors (2), (3), (4) or (5)) to any B-cell tumor subtype, namely HeH, 11q23/MLL,
828 t(1;19), t(9;22), dic(9;20) for ALL, C1 and C2 for MCL, n-CLL, i-CLL and m-CLL for CLL,

829 and finally GCB or ABC for DLBCLs. ALL, acute lymphoblastic leukemia; MCL, mantle cell
830 lymphoma; CLL, chronic lymphocytic leukemia; DLBCL, Diffuse large B cell lymphoma;
831 MM, multiple myeloma.

832 **b,** Accuracy for the pan-B-cell cancer diagnostic classifier (formed by the 5 predictors in
833 Fig. 3a) for training and validation series. Sensitivity is represented as black circles or
834 triangles for training or validation series, respectively (sensitivity in the training series
835 was evaluated using 10-fold stratified cross-validation). The total number of samples
836 used for both training and validation is shown.

837

838 **Fig. 4: Patient-specific DNA methylation changes are associated with silent chromatin**
839 **without an impact on gene expression.**

840 **a,** Circos plot representing the number of DNA methylation changes with respect to
841 heamatopoietic precursor cells (HPC) in individual patients for normal and neoplastic B
842 cells (each bar represents one patient). Total number of DNA methylation changes,
843 hypomethylation changes and hypermethylation changes are depicted at outer,
844 middle and inner tracks, respectively. Changes are further classified and color-coded as
845 B-cell related or B-cell independent if they occur or not during normal B-cell
846 differentiation, respectively.

847 **b,** Number of B-cell related changes against B-cell independent changes in normal and
848 neoplastic B-cells. Fitted regression lines are shown at bottom per each B-cell tumor
849 subtype.

850 **c,** B-cell related CpGs losing DNA methylation in B-cell tumors and the percentages in
851 each chromatin state in normal and neoplastic B-cells.

852 **d,** Example CpGs from **c** in normal and neoplastic B-cells.

853 **e,** Density of genes distributed along gene expression percentiles of genes associated
854 with B-cell related CpGs losing DNA methylation in B-cell tumors in each B-cell tumor
855 subtype. Expressed genes are displayed at right as control (with presence of
856 H3K36me3). Means within each B-cell subpopulation as well as B-cell tumors are
857 represented.
858 **f,** B-cell related CpGs gaining DNA methylation in B-cell tumors and the percentages in
859 each chromatin state in normal and neoplastic B-cells.
860 **g,** Example CpGs from **f** in normal and neoplastic B-cells.
861 **h,** Density of genes distributed along gene expression percentiles of genes associated
862 with B-cell related CpGs gaining DNA methylation in B-cell tumors in each B-cell tumor
863 subtype. Expressed genes are displayed at right as control (with presence of
864 H3K36me3). Means within each B-cell subpopulation as well as B-cell tumors are
865 represented.
866 **i,** Model for DNA methylation changes occurring at repressed regions during mitotic
867 cell division.
868 **j,** *In vitro* model for plasma blast differentiation from human primary naïve B cells.
869 Primary naïve B cells are labeled with CFSE at day 0. DNA methylation profiles are
870 obtained at day 0 (primary naïve B cells) day 4 (CFSE-high and CFSE-low) and day 6 (P1-
871 CFSE-low and CD38+, P2-CFSE-intermediate, CD38-, P3-CFSE-high CD38-).
872 **k,** DNA methylation changes accumulate at repressed regions during mitotic cell
873 division upon *in vitro* proliferation and differentiation of primary naïve B cells to
874 plasma blasts. Hypermethylation takes place at H3K27me3 regions, whereas
875 hypomethylation at heterochromatin and H3K9me3 regions. ChIP-seq data for primary

876 NBC and GC B cells was used. The mean of several biological replicates is represented
877 (numbers are depicted at bottom of the heatmaps).
878 **l,** Density of genes distributed in gene expression percentiles for genes with changes in
879 DNA methylation in (K).
880
881 **Fig. 5: Development and validation of an epigenetic mitotic clock reflecting the**
882 **proliferative history of normal and neoplastic B cells.**
883 **a,** Selection of CpGs gaining DNA methylation at H3K27me3 and losing DNA
884 methylation in normal and neoplastic B cells to construct the epiCMIT-hyper and
885 epiCMIT-hypo scores. epiCMIT, epigenetically-determined Cumulative MIToses.
886 **b,** Correlation of epiCMIT-hyper and epiCMIT-hypo in normal and neoplastic B cells.
887 **c,** epiCMIT in normal and neoplastic B cells.
888 **d,** Correlation of epiCMIT with previously reported hypermethylation-based mitotic
889 clocks, including epiTOC (epigenetic Timer for Cancer risk) and MiAge. Correlation with
890 CIMP (CpG hypermethylator phenotype) is also shown, although it has not been
891 formally presented as mitotic clock.
892 **e,** epiCMIT score during the *in vitro* B-cell differentiation as well as gene expression of
893 genes containing epiCMIT-CpGs. Expressed genes are also shown as control (presence
894 of H3K36me3).
895 **f,** Correlation of epiCMIT and mutational signatures related to proliferative history for
896 138 CLL samples with available DNA methylation data and WGS. SBS1, SBS5 have been
897 reported as clock-like mutational signatures. SBS9 is related to ncAID mutations.
898

899 **Fig. 6: The epiCMIT is a strong independent variable predicting clinical behavior in B-**
900 **cell tumors.**
901 **a,** epiCMIT is a dynamic variable that reflects disease progression. Precursor conditions
902 MGUS and MBL show significantly lower epiCMIT levels than their respective cancer
903 conditions CLL and MM, respectively. epiCMIT increases from diagnosis to progression
904 in paired CLL samples and in paired ALL at diagnosis, first relapse and second relapse.
905 **b,** Kaplan-Meier curves for three ALL cytogenetic groups shown as example divided in
906 low and high epiCMIT using maxstat rank statistics. Multivariate cox regression model
907 for relapse-free survival with epiCMIT against cytogenetic groups. Hazard ratio for
908 epiCMIT correspond to 0.1 increments.
909 **c,** Kaplan-Meier curves for CLL epigenetic groups based on different cellular origin
910 divided in low and high epiCMIT using maxstat rank statistics. Multivariate cox
911 regression model for time to first treatment with epiCMIT against age, number of
912 driver alterations and epigenetic groups based on different cellular origin. Validation
913 series is shown in **d**. Hazard ratio for epiCMIT correspond to 0.1 increments.
914 **e,** Kaplan-Meier curves for MCL epigenetic groups based on different cellular origin
915 divided in low and high epiCMIT using maxstat rank statistics. Multivariate cox
916 regression model for overall survival with epiCMIT against epigenetic groups based on
917 different cellular origin. Hazard ratio for epiCMIT correspond to 0.1 increments.
918 **f,** Validation series for C1 MCL. Hazard ratio for epiCMIT correspond to 0.1 increments.
919
920 **Fig. 7: epiCMIT is associated with specific genetic driver alterations in CLL**
921 **a,** Are there specific genetic alterations in CLL that may predispose a proliferative
922 advantage to neoplastic cells?

923 **b,** Genetic driver alterations in CLL associated with higher epiCMIT grouped by
924 signaling pathways reported. Analyses were done globally for all CLL samples (although
925 adjusted by epigenetic groups) as well as within each epigenetic subgroup. Positive
926 point estimates relate to positive associations with epiCMIT. 95% confidence intervals
927 are shown with colors coding for FDR corrections. Number of patients with each
928 genetic driver alteration is shown at right.
929 **c,** Oncoprint showing genetic driver alterations associated with higher epiCMIT in CLL
930 epigenetic groups separately. Other clinicobiological features including MBL or CLL,
931 IGHV status, Age, Binet stage, epiCMIT subgroups based on Fig. 5c, need for treatment
932 and patient status are shown. Cases are ordered within each CLL subgroup from lower
933 to higher epiCMIT values. Distinct genetic driver alterations are depicted with different
934 colors and shapes. The percentage of mutated patients as well as barplots showing the
935 number of mutated patients for each alteration is shown at right.
936
937
938

939

940 **Supplementary figure legends**

941

942 **Extended Data Fig.1**

943 **a,** Principal component analysis and hierarchical clustering of paired un/purified DNA
944 methylation profiles obtained with EPIC array from MCL and CLL patients. Colors
945 represent the same un/purified sample, with FACS-based purities highlighted in each
946 sample. MCL, mantle cell lymphoma. CLL, chronic lymphocytic leukemia.

947 **b,** Correlations and Passing Bablock regression fits of gold-standard methods for tumor
948 purity prediction (FACS and genetic-based) against DNA methylation-based tumor
949 purity prediction for MCL and CLL patients.

950 **c,** Pearson correlations and Passing Bablock regression fits for gold-standard methods
951 for tumor purity predictions (FACS and genetic-based) against DNA methylation-based
952 tumor purity predictions for MM and DLBCL patients. MM, multiple myeloma. DLBCL,
953 Diffuse large B cell lymphoma.

954 **d,** Pan-B cell DNA methylation signature used to deconvolute DNA methylation and
955 obtain B-cell tumor purities in DLBCL and MM. Bar plots representing DNA-methylation
956 based predictions as well as gold standard-based predictions are represented at top
957 the heatmaps.

958 **e,** Genomic distribution of stably un/methylated CpGs in B-cell tumors.

959 **f,** Example gene showing stably un/methylated CpGs.

960 **g,** Gene ontology analysis of genes showing both stable un/methylated CpGs.

961

962

963 **Extended Data Fig.2**

964 **a,** Principal component analysis for normal and neoplastic B cells, including ALL, MCL,
965 CLL, DLBCL and MM. Each component is resented separately (until the ninth
966 component). ALL, acute lymphoblastic leukemia; MCL, mantle cell lymphoma; CLL,
967 chronic lymphocytic leukemia; DLBCL, Diffuse large B cell lymphoma; MM, multiple
968 myeloma.

969 **b,** Heatmap showing B-cell tumor-specific hypermethylation and the number of CpGs
970 falling at regulatory regions. Per each B-cell tumor, the same number of *de novo* CpGs
971 was randomly chosen from the 450K array and interrogated the percentage falling at
972 regulatory regions.

973 **c,** Genomic distribution for *de novo* DNA methylation changes in B-cell tumors.

974 **d,** Gene expression of DNA methyltransferases *DNMT1*, *DNMT3A* and *DNMT3B*
975 throughout B cell differentiation.

976 **e,** Gene expression percentile of TFs showing the most significant p-values and
977 frequencies in *de novo* hypomethylation signatures in each B-cell tumor in Fig. 2d.

978

979 **Extended Data Fig. 3**

980 **a,** Sensitivity of the pan-B-cell diagnostic algorithm for the classification of an unknown
981 B-cell tumor into ALL, MCL, CLL, DLBCL or MM while incrementing the number of CpGs
982 used in the classifier algorithm (first step of Fig. 3A). The number of CpGs selected for
983 the predictor is selected by maximizing the highest balanced accuracy and is indicated
984 with a red circle. This strategy was applied also in the remaining 4 predictors to classify
985 B-cell tumor subtypes in panels b,, c,, d, and e, (second step of Fig. 3A). Each B-cell
986 tumor is represented with different shapes and colors. ALL, acute lymphoblastic

987 leukemia; MCL, mantle cell lymphoma; CLL, chronic lymphocytic leukemia; DLBCL,
988 Diffuse large B cell lymphoma; MM, multiple myeloma
989 **b,** Sensitivity of predictor 2 for the pan-B-cell diagnostic algorithm (predictor 2 of Fig.
990 3a) for the classification of ALL into the subtypes HeH, 11q23/MLL, t(12;21), t(1;19),
991 t(9;22) and dic(9;20) while incrementing the number of CpGs.
992 **c,** Sensitivity of predictor 3 of the pan-B-cell diagnostic algorithm (predictor 3 of Fig.
993 3a) for the classification of MCL into the subtypes C1 or C2 while incrementing the
994 number of CpGs.
995 **d,** Sensitivity of predictor 4 of the pan-B-cell diagnostic algorithm (predictor 4 of Fig.
996 3a) for the classification of CLL into the subtypes n-CLL, i-CLL or m-CLL while
997 incrementing the number of CpGs.
998 **e,** Sensitivity of predictor 5 of the pan-B-cell diagnostic algorithm (predictor 5 of Fig.
999 3a) for the classification of DLBCL into the subtypes ABC and GCB while incrementing
1000 the number of CpGs.
1001
1002 **Extended Data Fig.4**
1003 **a,** Variability of DNA methylation changes (IQR) in normal neoplastic B cells against the
1004 median number of DNA methylation changes.
1005 **b,** Correlations of B-cell independent DNA methylation changes in B-cell tumors
1006 against B-cell related changes for hypermethylation (top) and hypomethylation
1007 (bottom).
1008 **c,** Number of B-cell related or B-cell independent hyper- or hypomethylation in B-cell
1009 tumors (Methods).

1010 **d,** B-cell independent CpGs losing DNA methylation in B-cell tumors and the
1011 percentages in each chromatin state in normal and neoplastic B-cells.
1012 **e,** Example CpGs from **d,** in normal and neoplastic B-cells.
1013 **f,** Density of genes distributed along gene expression percentiles of genes associated
1014 with B-cell independent CpGs losing DNA methylation in B-cell tumors in each B-cell
1015 tumor subtype. Expressed genes are displayed at right as control (with presence of
1016 H3K36me3). Means within each B-cell subpopulation as well as B-cell tumors are
1017 represented.
1018 **g,** B-cell independent CpGs gaining DNA methylation in B-cell tumors and the
1019 percentages in each chromatin state in normal and neoplastic B-cells.
1020 **h,** Example CpGs from g, in normal and neoplastic B-cells.
1021 **i,** Density of genes distributed along gene expression percentiles of genes associated
1022 with B-cell independent CpGs gaining DNA methylation in B-cell tumors in each B-cell
1023 tumor subtype. Expressed genes are displayed at right as control (with presence of
1024 H3K36me3). Means within each B-cell subpopulation as well as B-cell tumors are
1025 represented.
1026
1027 **Extended Data Fig.5.**
1028 **a,** Correlations between epiCMIT-hyper and epiCMIT-hypo with other
1029 hypermethylation-based mitotic clocks including epiTOC (epigenetic Timer for Cancer
1030 risk) and MiAge as well as the CIMP (CpG hypermethylator phenotype, not formally
1031 conceived as mitotic clock). epiCMIT, epigenetically-determined Cumulative MIToses.

1032 **b,** Poor effect of age in epiCMIT in normal B cells. epiCMIT correlation with age in naïve
1033 B cells from healthy infants, young-adults and older adults. Horvath model correctly
1034 predicted the real age. epiCMIT and Horvath are poorly correlated.
1035 **c,** Poor effect of age in epiCMIT in B-ALL. Wide epiCMIT spectrum in children with B-
1036 ALL.
1037 **d,** Correlation between the number of mutations detected by WGS against epiCMIT in
1038 CLL subtypes with different cellular origin (n-CLL, i-CLL and m-CLL).
1039 **e,** Mutational signatures for 138 CLL samples with available WGS and DNA methylation
1040 data divided into different subtypes of CLL with different cellular origin, namely n-CLL,
1041 i-CLL and m-CLL.
1042 **f,** Correlation of epiCMIT with Ki67 in nodal MCL patients.
1043 **g,** Gene expression signatures resulting after GSEA analysis related to proliferation in
1044 CLL samples with low and high epiCMIT.
1045 **h,** Gene expression signatures resulting after GSEA analysis related to proliferation in
1046 ALL samples with low and high epiCMIT.
1047
1048 **Extended Data Fig.6.**
1049 **a,** epiCMIT interpretation in normal and neoplastic B cells. Mitotic cell division occurs
1050 in normal B cell development and DNA methylation changes accumulate at repressed
1051 regions (blue component of the epiCMIT on the x axis). B-cell tumors arise from
1052 different maturation stages, and thus they contain different baseline epiCMIT from the
1053 cell of origin from which they originate (blue bar in B-cell tumors). When B-cell tumors
1054 progress, they acquire additional DNA methylation changes at repressed regions
1055 (shown as red component of the epiCMIT). Notably, these new DNA methylation

1056 changes occur in both axes, i.e. changes occurring in normal B cell maturation as well
1057 as new B-cell independent changes. This concept is shown with real data at Fig. 4b.
1058 **b,** Multivariate cox regression model for overall survival in ALL for epiCMIT against
1059 cytogenetic groups. Hazard ratio for epiCMIT correspond to 0.1 increments.
1060 **c,** Multivariate cox regression model for time to first treatment in CLL samples near
1061 diagnosis (30 months) for epiCMIT against age, number of driver alterations and
1062 epigenetic groups based on different cellular origin. At right is shown the multivariate
1063 cox regression model for validation series. Hazard ratio for epiCMIT correspond to 0.1
1064 increments.
1065 **d,** Multivariate cox regression model in CLL for overall survival for epiCMIT against age,
1066 number of driver alterations and epigenetic groups for the initial and validation series
1067 (left and right, respectively). Treated and untreated patients were included in the
1068 validation series. Hazard ratio for epiCMIT correspond to 0.1 increments.
1069 **e,** Kaplan-Meyer curves for DLBCL separated by ABC and GCB groups and low and high
1070 epiCMIT groups based on maxstat rank statistics. Multivariate cox regression model
1071 treating epiCMIT as continuous variable is shown at bottom. Hazard ratio for epiCMIT
1072 correspond to 0.1 increments.
1073
1074 **Extended Data Fig.7.**
1075 **a,** Oncoprint showing all genetic driver alterations considered in CLL grouped by
1076 epigenetic subgroups and ordered according to increasing levels of epiCMIT (from left
1077 to right within each epigenetic subgroup). Other clinicobiological features including
1078 MBL or CLL, IGHV status, Age, Binet stage, epiCMIT subgroups based on Fig. 5C, need
1079 for treatment and patient status are shown. Distinct genetic driver alterations are

1080 depicted with different colors and shapes. The percentage of mutated patients as well
1081 as barplots showing the number of mutated patients for each alteration is shown at
1082 right.
1083 **b,** Driver genetic alterations without clear associations with epiCMIT. Analyses were
1084 done globally for all CLL samples (although adjusted by epigenetic groups) as well as
1085 within each epigenetic subgroup. Positive point estimates relate to positive
1086 associations with epiCMIT. 95% confidence intervals are shown with colors coding for
1087 FDR corrections. Number of patients with each genetic driver alteration is shown at
1088 bottom.
1089

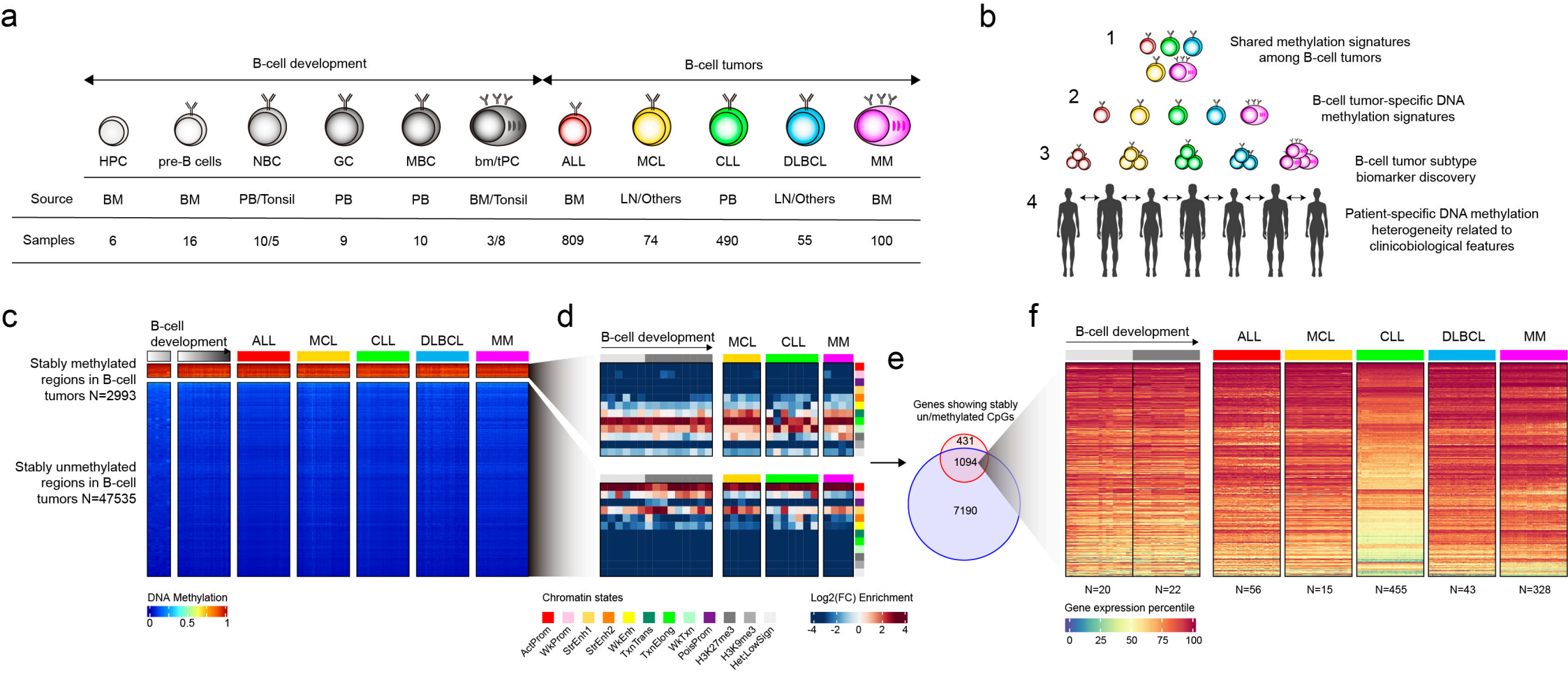

a
B-cell development
B-cell tumors
HPC
pre-B cells
NBC
GC
MBC
bm/tPC
ALL
MCL
CLL
DLBCL
MM
Source
BM
BM
PB/Tonsil
PB
PB
BM/Tonsil
BM
LN/Others
PB
LN/Others
BM
Samples
6
16
10/5
9
10
3/8
809
74
490
55
100
b
1
Shared methylation signatures among B-cell tumors
2
B-cell tumor-specific DNA methylation signatures
3
B-cell tumor subtype biomarker discovery
4
Patient-specific DNA methylation heterogeneity related to clinicobiological features
c
B-cell development
ALL
MCL
CLL
DLBCL
MM
Stably methylated regions in B-cell tumors N=2993
Stably unmethylated regions in B-cell tumors N=47535
DNA Methylation
0
0.5
1
d
B-cell development
MCL
CLL
MM
Chromatin states
ActProm
WkProm
StrEnh1
StrEnh2
WkEnh
TxnTrans
TxnElong
WkTxn
PoisProm
H3K27me3
H3K9me3
HetLowSign
Log2(FC) Enrichment
-4
-2
0
2
4
e
Genes showing stably un/methylated CpGs
431
1094
7190
f
B-cell development
ALL
MCL
CLL
DLBCL
MM
N=20
N=22
N=56
N=15
N=455
N=43
N=328
Gene expression percentile
0
25
50
75
100

Fig. 1

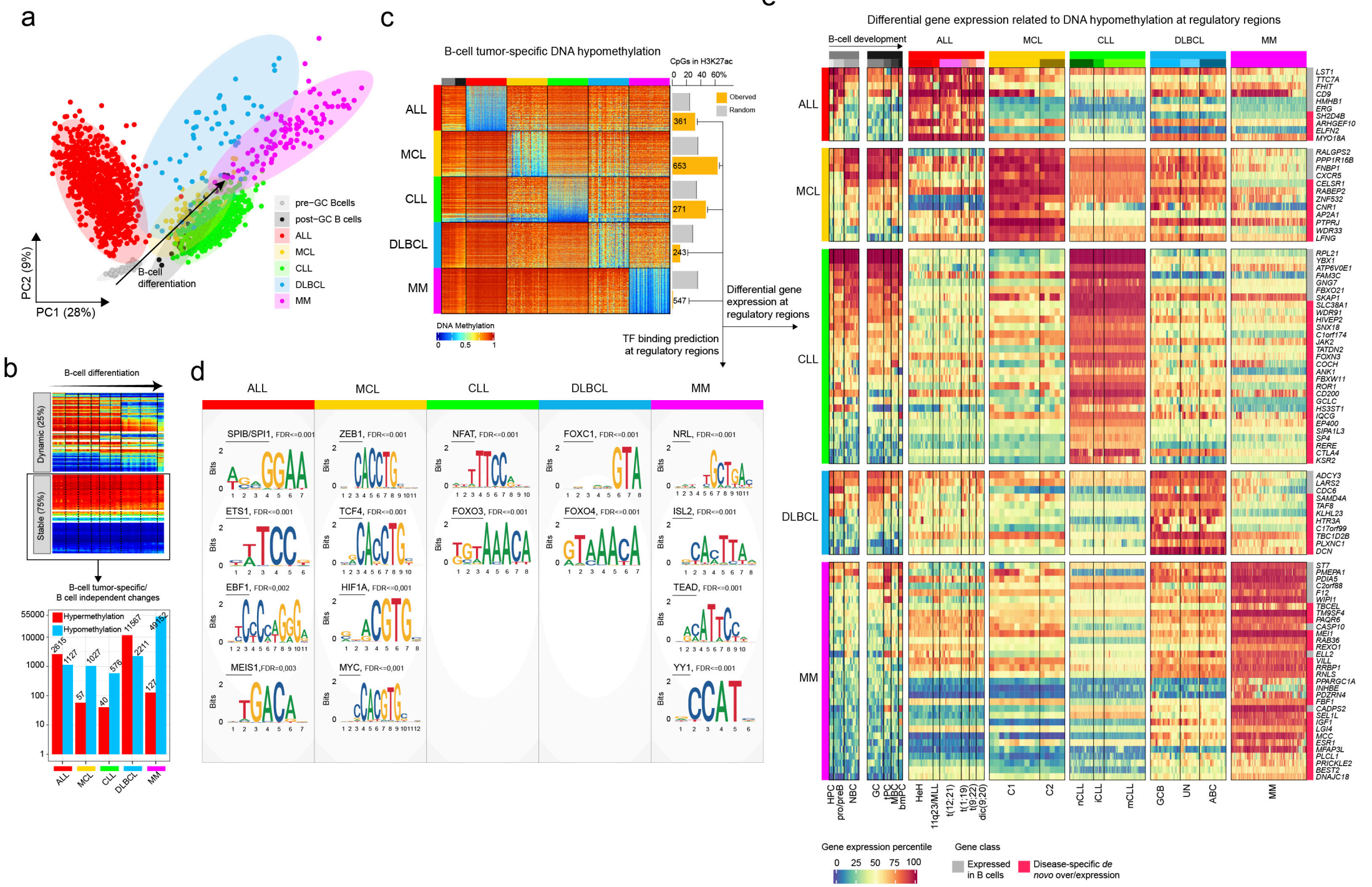

a
PC2 (9%)
PC1 (28%)
B-cell differentiation
pre−GC Bcells
post−GC B cells
ALL
MCL
CLL
DLBCL
MM
b
B-cell differentiation
Dynamic (25%)
Stable (75%)
B-cell tumor-specific/
B cell independent changes
Hypermethylation
Hypomethylation
c
B-cell tumor-specific DNA hypomethylation
CpGs in H3K27ac
Oberved
Random
361
653
271
243
547
DNA Methylation
Differential gene expression at regulatory regions
TF binding prediction at regulatory regions
d
SPIB/SPI1, FDR<=0.001
ZEB1, FDR<=0.001
NFAT, FDR<=0.001
FOXC1, FDR<=0.001
NRL, FDR<=0.001
ETS1, FDR<=0.001
TCF4, FDR<=0.001
FOXO3, FDR<=0.001
FOXO4, FDR<=0.001
ISL2, FDR<=0.001
EBF1, FDR=0,002
HIF1A, FDR<=0,001
TEAD, FDR<=0.001
MEIS1,FDR=0,003
MYC, FDR<=0,001
YY1, FDR<=0.001
e
Differential gene expression related to DNA hypomethylation at regulatory regions
B-cell development
Gene expression percentile
Gene class
Expressed in B cells
Disease-specific de novo over/expression

Fig. 2

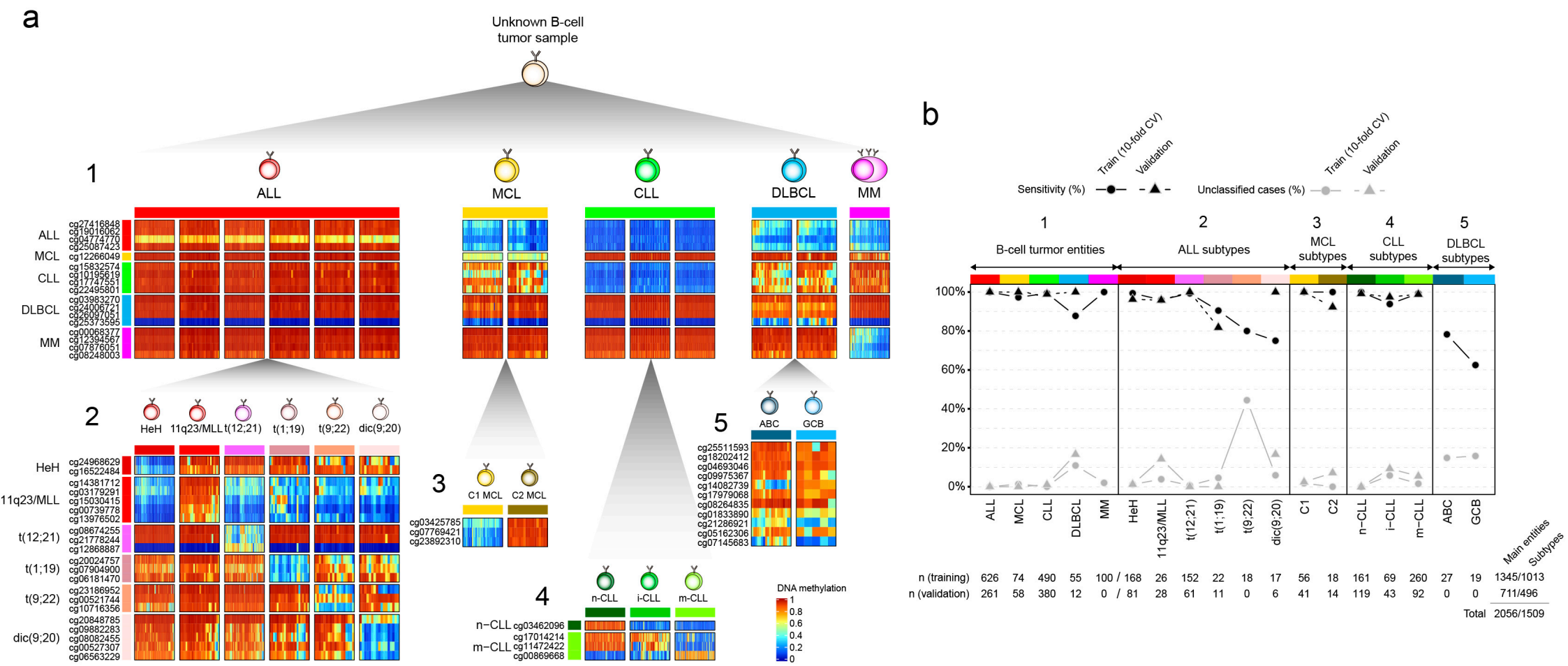

a
Unknown B-cell tumor sample
1
ALL
MCL
CLL
DLBCL
MM
2
HeH
11q23/MLL
t(12;21)
t(1;19)
t(9;22)
dic(9;20)
3
C1 MCL
C2 MCL
4
n-CLL
i-CLL
m-CLL
5
ABC
GCB
DNA methylation
b
Train (10-fold CV)
Validation
Sensitivity (%)
Unclassified cases (%)
1
B-cell turmor entities
2
ALL subtypes
3
MCL subtypes
4
CLL subtypes
5
DLBCL subtypes
100%
80%
60%
40%
20%
0%
ALL
MCL
CLL
DLBCL
MM
HeH
11q23/MLL
t(12;21)
t(1;19)
t(9;22)
dic(9;20)
C1
C2
n−CLL
i−CLL
m−CLL
ABC
GCB
Main entities
Subtypes
n (training) 626 74 490 55 100 / 168 26 152 22 18 17 56 18 161 69 260 27 19 1345/1013
n (validation) 261 58 380 12 0 / 81 28 61 11 0 6 41 14 119 43 92 0 0 711/496
Total 2056/1509

Fig. 3

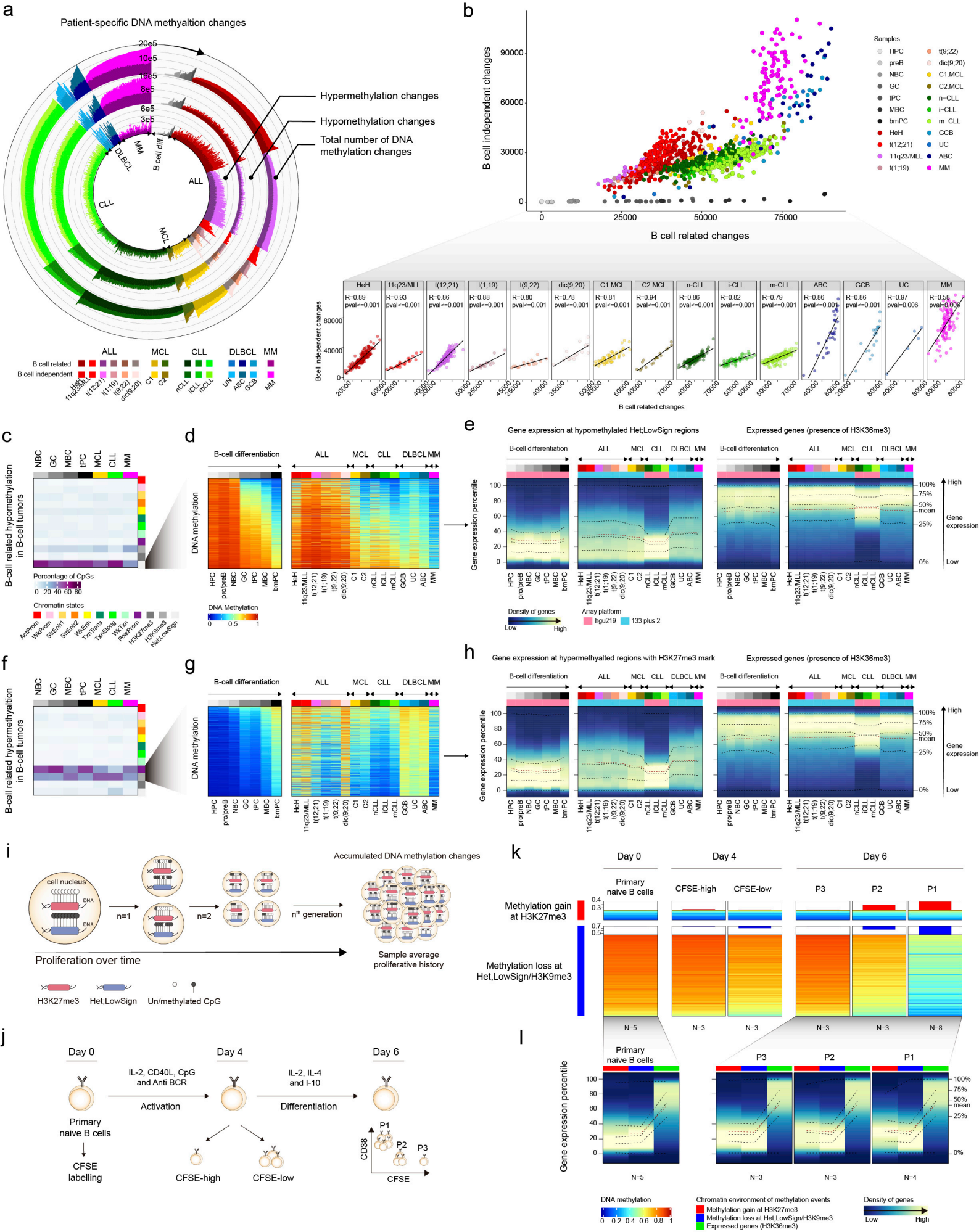

a
Patient-specific DNA methyaltion changes
Hypermethylation changes
Hypomethylation changes
Total number of DNA methylation changes
B cell related
B cell independent
b
B cell independent changes
B cell related changes
Samples
c
B-cell related hypomethylation in B-cell tumors
Percentage of CpGs
Chromatin states
d
B-cell differentiation
DNA methylation
DNA Methylation
e
Gene expression at hypomethylated Het;LowSign regions
Expressed genes (presence of H3K36me3)
Gene expression percentile
Density of genes
Array platform
hgu219
133 plus 2
f
B-cell related hypermethylation in B-cell tumors
g
h
Gene expression at hypermethyalted regions with H3K27me3 mark
i
cell nucleus
Proliferation over time
Accumulated DNA methylation changes
Sample average proliferative history
H3K27me3
Het;LowSign
Un/methylated CpG
j
Day 0
Day 4
Day 6
IL-2, CD40L, CpG and Anti BCR
Activation
IL-2, IL-4 and I-10
Differentiation
Primary naive B cells
CFSE labelling
CFSE-high
CFSE-low
CD38
CFSE
k
Methylation gain at H3K27me3
Methylation loss at Het,LowSign/H3K9me3
l
DNA methylation
Chromatin environment of methylation events
Methylation gain at H3K27me3
Methylation loss at Het;LowSign/H3K9me3
Expressed genes (H3K36me3)
Density of genes

Fig. 4

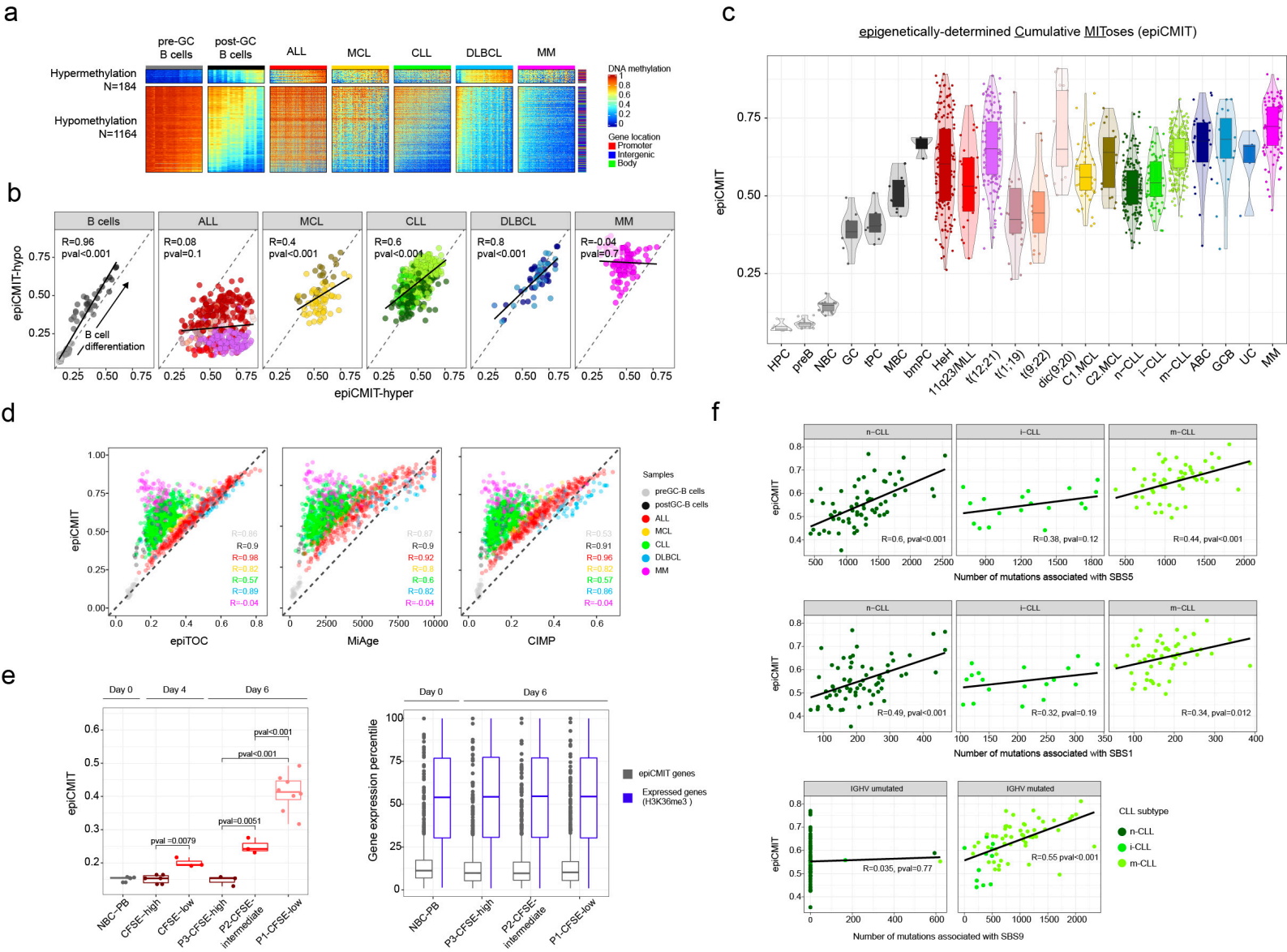

a
pre-GC B cells
post-GC B cells
ALL
MCL
CLL
DLBCL
MM
Hypermethylation N=184
Hypomethylation N=1164
DNA methylation
Gene location
Promoter
Intergenic
Body
b
B cells
ALL
MCL
CLL
DLBCL
MM
R=0.96 pval<0.001
R=0.08 pval=0.1
R=0.4 pval<0.001
R=0.6 pval<0.001
R=0.8 pval<0.001
R=-0.04 pval=0.7
B cell differentiation
epiCMIT-hypo
epiCMIT-hyper
c
epigenetically-determined Cumulative MIToses (epiCMIT)
epiCMIT
HPC
preB
NBC
GC
tPC
MBC
bmPC
HeH
11q23/MLL
t(12;21)
t(1;19)
t(9;22)
dic(9;20)
C1.MCL
C2.MCL
n-CLL
i-CLL
m-CLL
ABC
GCB
UC
MM
d
epiCMIT
epiTOC
MiAge
CIMP
Samples
preGC-B cells
postGC-B cells
ALL
MCL
CLL
DLBCL
MM
e
Day 0
Day 4
Day 6
epiCMIT
pval =0.0079
pval=0.0051
pval<0.001
NBC-PB
CFSE-high
CFSE-low
P3-CFSE-high
P2-CFSE-intermediate
P1-CFSE-low
Gene expression percentile
epiCMIT genes
Expressed genes (H3K36me3)
f
n-CLL
i-CLL
m-CLL
R=0.6, pval<0.001
R=0.38, pval=0.12
R=0.44, pval<0.001
Number of mutations associated with SBS5
R=0.49, pval<0.001
R=0.32, pval=0.19
R=0.34, pval=0.012
Number of mutations associated with SBS1
IGHV umutated
IGHV mutated
R=0.035, pval=0.77
R=0.55 pval<0.001
Number of mutations associated with SBS9
CLL subtype
n-CLL
i-CLL
m-CLL

Fig. 5

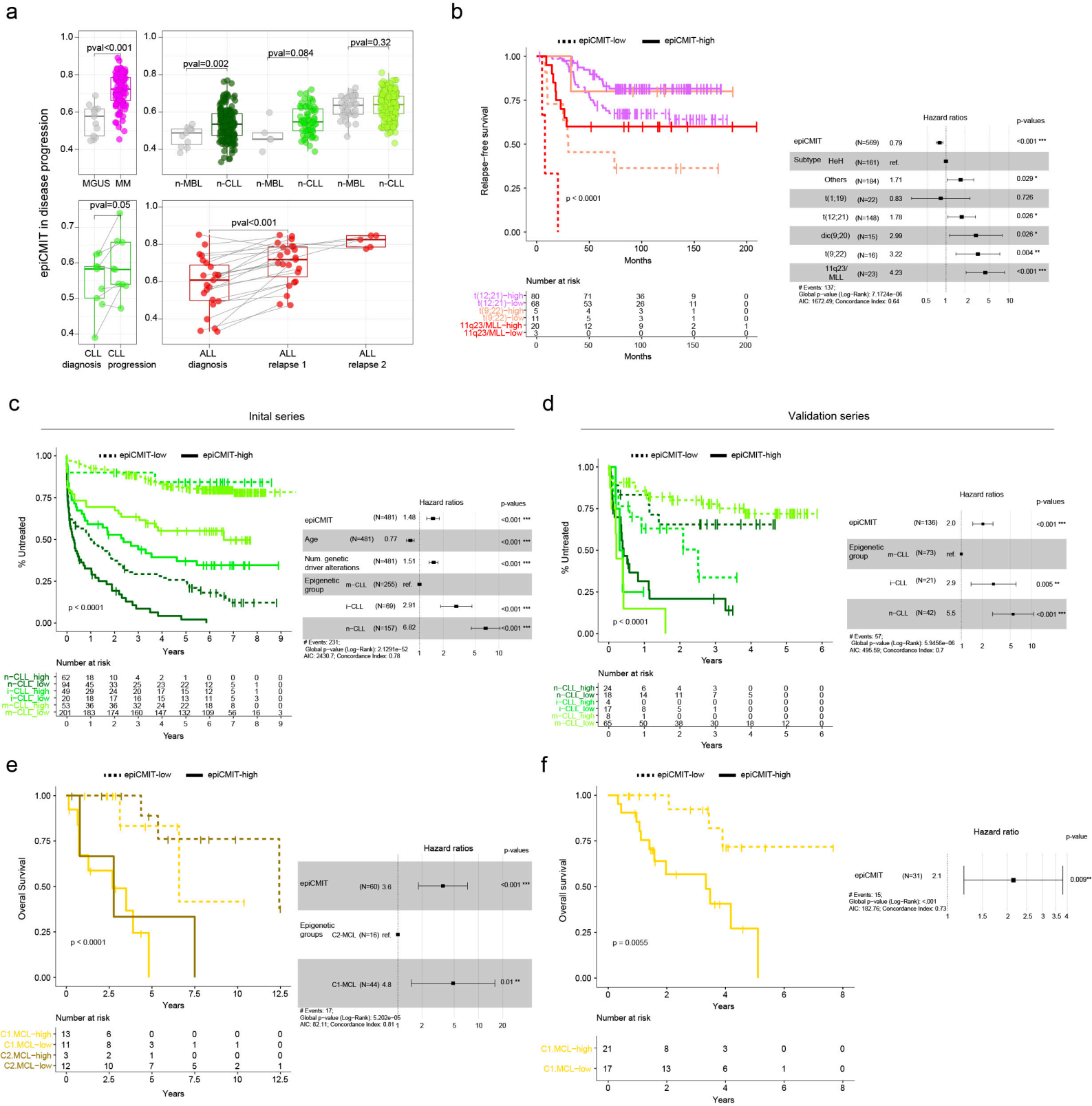

a
epiCMIT in disease progression
pval<0.001
MGUS MM
pval=0.002
pval=0.084
pval=0.32
n-MBL n-CLL n-MBL n-CLL n-MBL n-CLL
pval=0.05
CLL diagnosis
CLL progression
pval<0.001
ALL diagnosis
ALL relapse 1
ALL relapse 2
b
epiCMIT-low
epiCMIT-high
Relapse-free survival
p < 0.0001
Months
Number at risk
t(12;21)-high 80 71 36 9 0
t(12;21)-low 68 53 26 11 0
t(9;22)-high 5 4 3 1 0
t(9;22)-low 11 5 3 1 0
11q23/MLL-high 20 12 9 2 1
11q23/MLL-low 3 0 0 0 0
Hazard ratios
p-values
epiCMIT (N=569) 0.79 <0.001 ***
Subtype HeH (N=161) ref.
Others (N=184) 1.71 0.029 *
t(1;19) (N=22) 0.83 0.726
t(12;21) (N=148) 1.78 0.026 *
dic(9;20) (N=15) 2.99 0.026 *
t(9;22) (N=16) 3.22 0.004 **
11q23/MLL (N=23) 4.23 <0.001 ***
Events: 137;
Global p-value (Log-Rank): 7.1724e-06
AIC: 1672.49; Concordance Index: 0.64
c
Inital series
epiCMIT-low
epiCMIT-high
% Untreated
p < 0.0001
Years
Number at risk
n-CLL_high 62 18 10 4 2 1 0 0 0 0
n-CLL_low 94 45 33 25 23 22 12 5 1 0
i-CLL_high 49 29 24 20 17 15 12 5 1 0
i-CLL_low 20 18 17 16 15 13 11 5 3 0
m-CLL_high 53 36 36 32 24 22 18 8 0 0
m-CLL_low 201 183 174 160 147 132 109 56 16 3
Hazard ratios
p-values
epiCMIT (N=481) 1.48 <0.001 ***
Age (N=481) 0.77 <0.001 ***
Num. genetic driver alterations (N=481) 1.51 <0.001 ***
Epigenetic group m-CLL (N=255) ref.
i-CLL (N=69) 2.91 <0.001 ***
n-CLL (N=157) 6.82 <0.001 ***
Events: 231;
Global p-value (Log-Rank): 2.1291e-52
AIC: 2430.7; Concordance Index: 0.78
d
Validation series
epiCMIT-low
epiCMIT-high
% Untreated
p < 0.0001
Years
Number at risk
n-CLL_high 24 6 4 3 0 0 0
n-CLL_low 18 14 11 7 5 0 0
i-CLL_high 4 0 0 0 0 0 0
i-CLL_low 17 8 5 1 0 0 0
m-CLL_high 8 1 0 0 0 0 0
m-CLL_low 65 50 38 30 18 12 0
Hazard ratios
p-values
epiCMIT (N=136) 2.0 <0.001 ***
Epigenetic group m-CLL (N=73) ref.
i-CLL (N=21) 2.9 0.005 **
n-CLL (N=42) 5.5 <0.001 ***
Events: 57;
Global p-value (Log-Rank): 5.9456e-06
AIC: 495.59; Concordance Index: 0.7
e
epiCMIT-low
epiCMIT-high
Overal Survival
p < 0.0001
Years
Number at risk
C1.MCL-high 13 6 0 0 0 0
C1.MCL-low 11 8 3 1 1 0
C2.MCL-high 3 2 1 0 0 0
C2.MCL-low 12 10 7 5 2 1
Hazard ratios
p-values
epiCMIT (N=60) 3.6 <0.001 ***
Epigenetic groups C2-MCL (N=16) ref.
C1-MCL (N=44) 4.8 0.01 **
Events: 17;
Global p-value (Log-Rank): 5.202e-05
AIC: 82.11; Concordance Index: 0.81
f
epiCMIT-low
epiCMIT-high
Overall survival
p = 0.0055
Years
Number at risk
C1.MCL-high 21 8 3 0 0
C1.MCL-low 17 13 6 1 0
Hazard ratio
p-value
epiCMIT (N=31) 2.1 0.009**
Events: 15;
Global p-value (Log-Rank): <.001
AIC: 182.76; Concordance Index: 0.73

Fig. 6

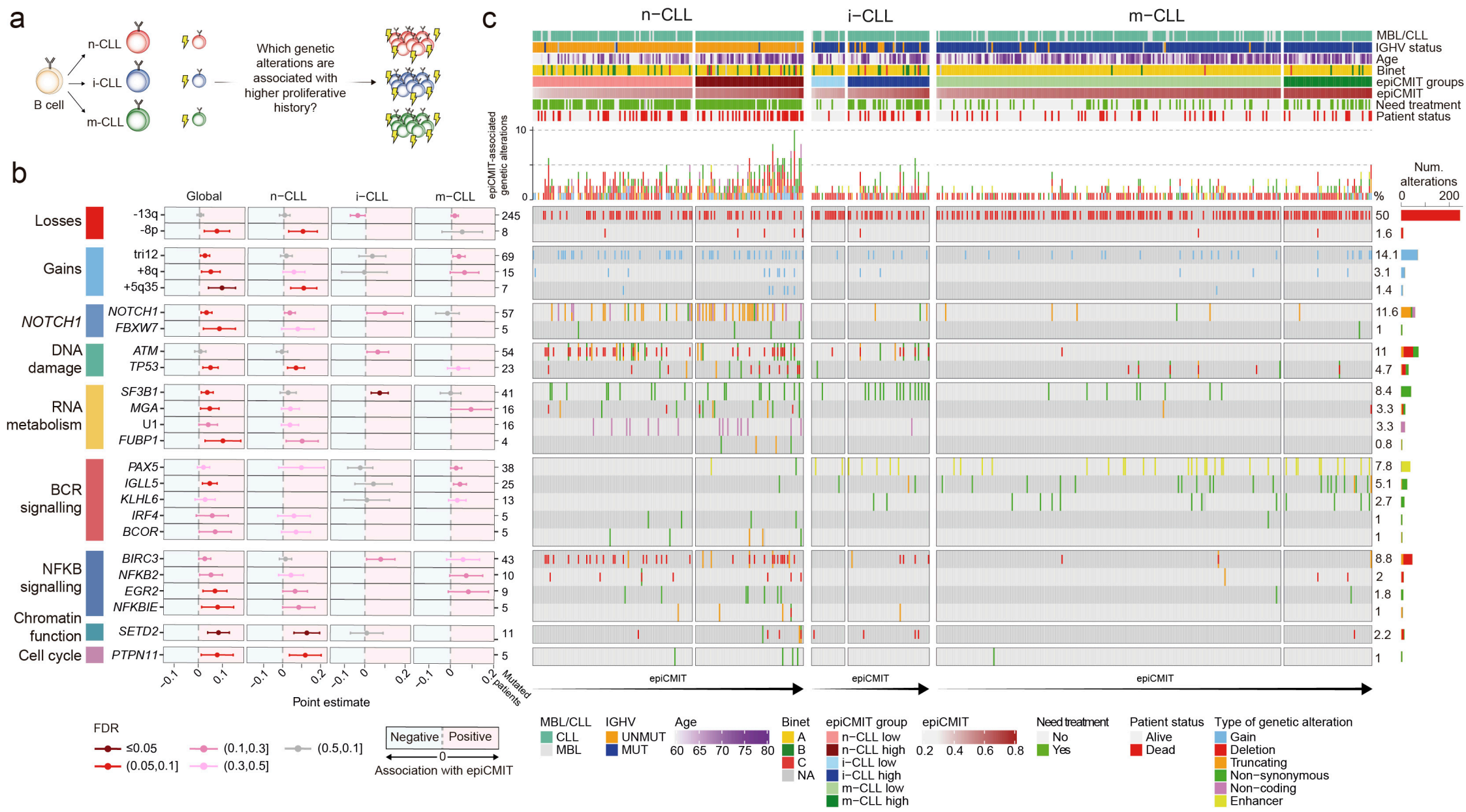
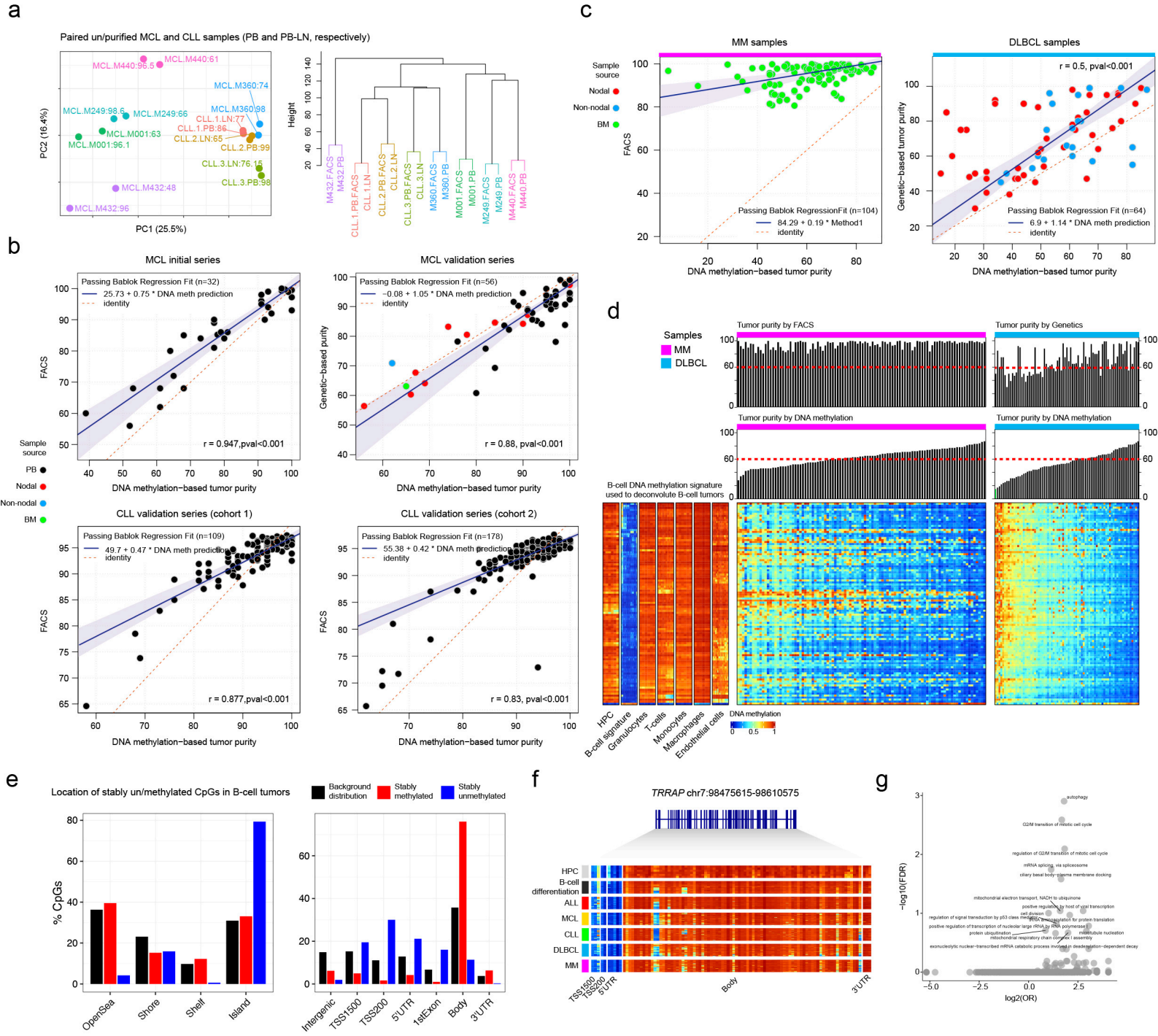
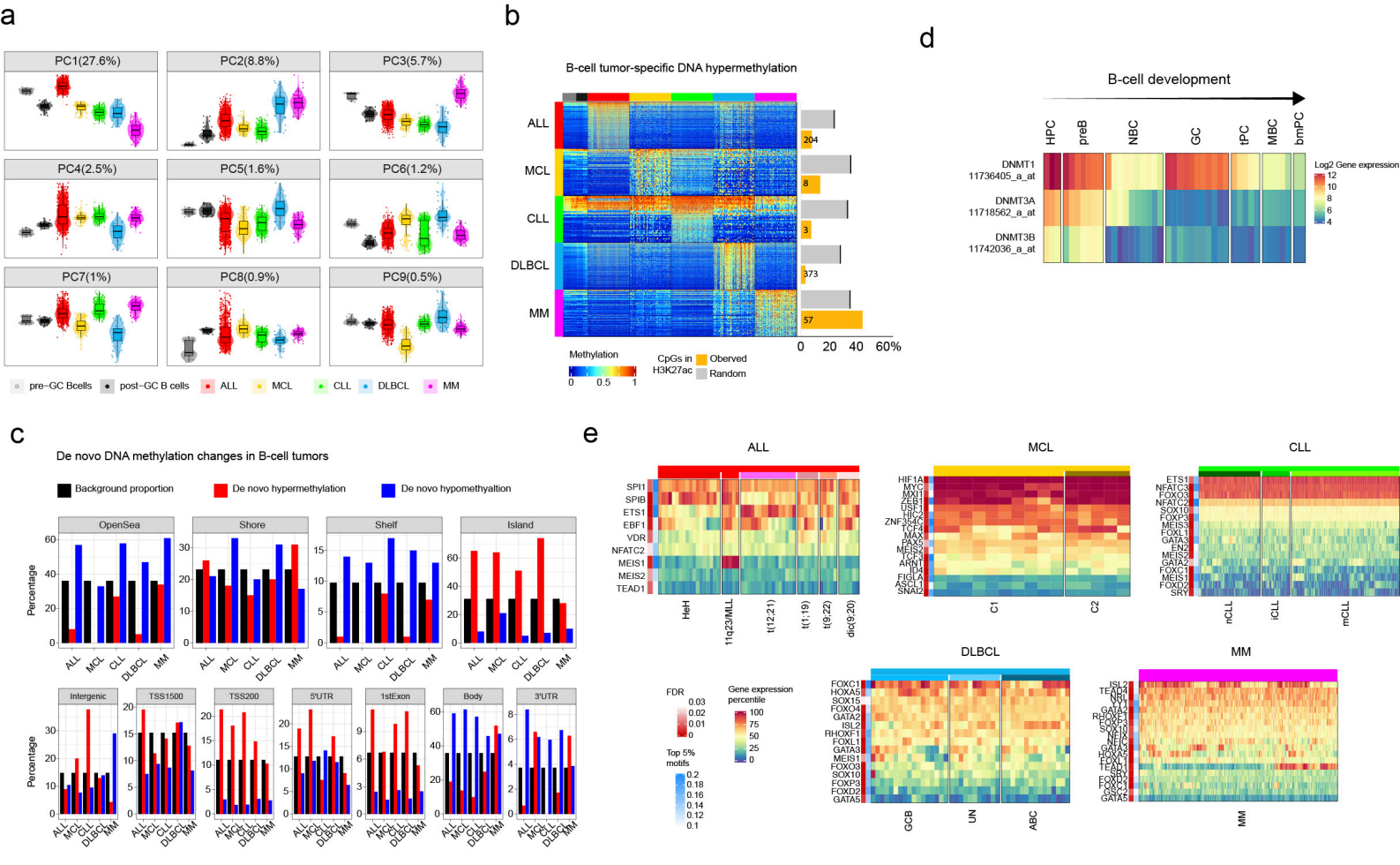

a
B cell
n-CLL
i-CLL
m-CLL
Which genetic alterations are associated with higher proliferative history?
b
Global
n-CLL
i-CLL
m-CLL
Losses
-13q
-8p
Gains
tri12
+8q
+5q35
NOTCH1
NOTCH1
FBXW7
DNA damage
ATM
TP53
RNA metabolism
SF3B1
MGA
U1
FUBP1
BCR signalling
PAX5
IGLL5
KLHL6
IRF4
BCOR
NFKB signalling
BIRC3
NFKB2
EGR2
NFKBIE
Chromatin function
SETD2
Cell cycle
PTPN11
Point estimate
Mutated patients
FDR
≤0.05
(0.05,0.1]
(0.1,0.3]
(0.3,0.5]
(0.5,0.1]
Negative
Positive
Association with epiCMIT
c
n-CLL
i-CLL
m-CLL
MBL/CLL
IGHV status
Age
Binet
epiCMIT groups
epiCMIT
Need treatment
Patient status
epiCMIT-associated genetic alterations
Num. alterations
epiCMIT
MBL/CLL
CLL
MBL
IGHV
UNMUT
MUT
Age
Binet
A
B
C
NA
epiCMIT group
n-CLL low
n-CLL high
i-CLL low
i-CLL high
m-CLL low
m-CLL high
epiCMIT
Need treatment
No
Yes
Patient status
Alive
Dead
Type of genetic alteration
Gain
Deletion
Truncating
Non-synonymous
Non-coding
Enhancer

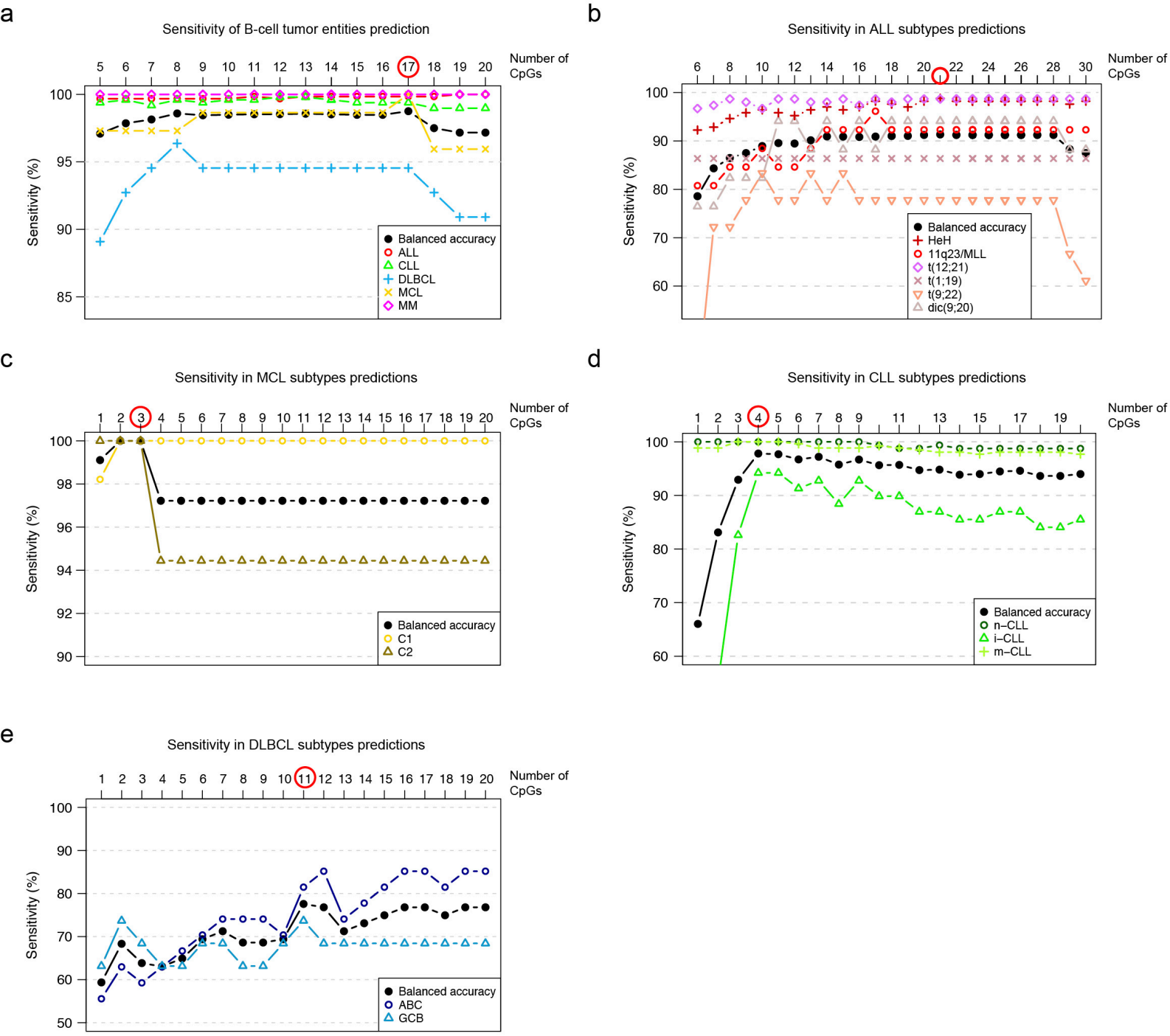
Fig. 7



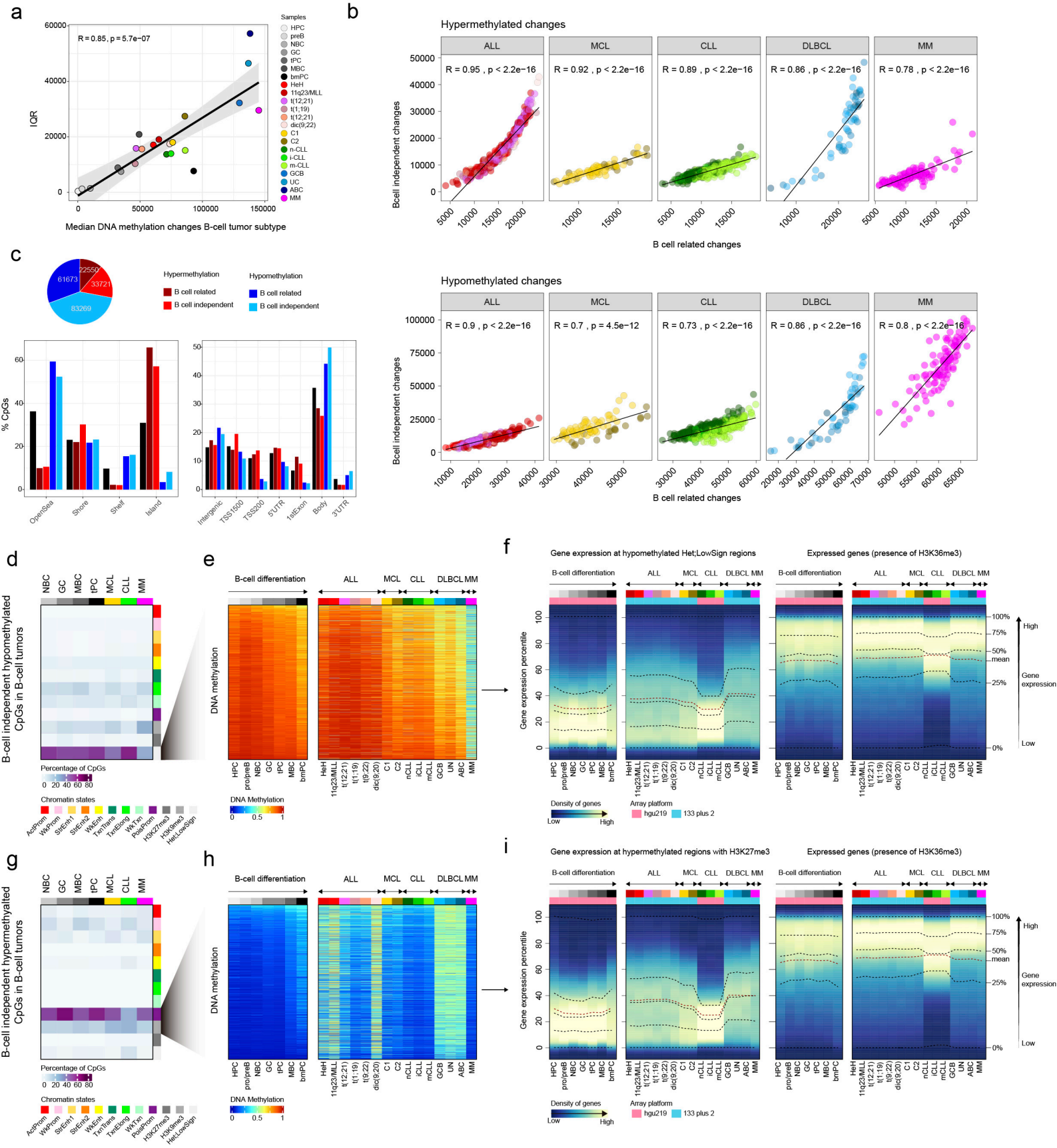
Extended Data Fig.1



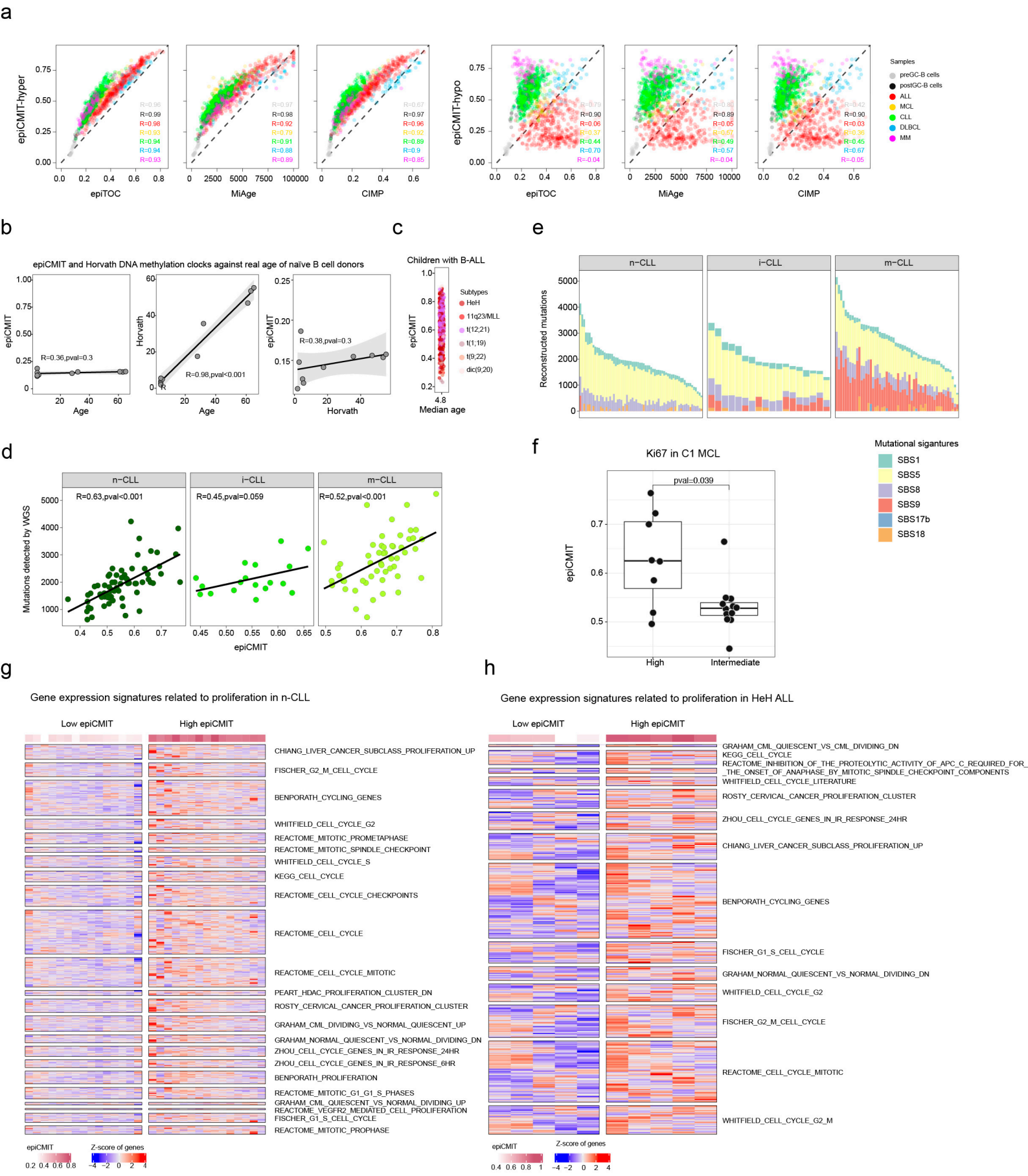
Extended Data Fig. 2



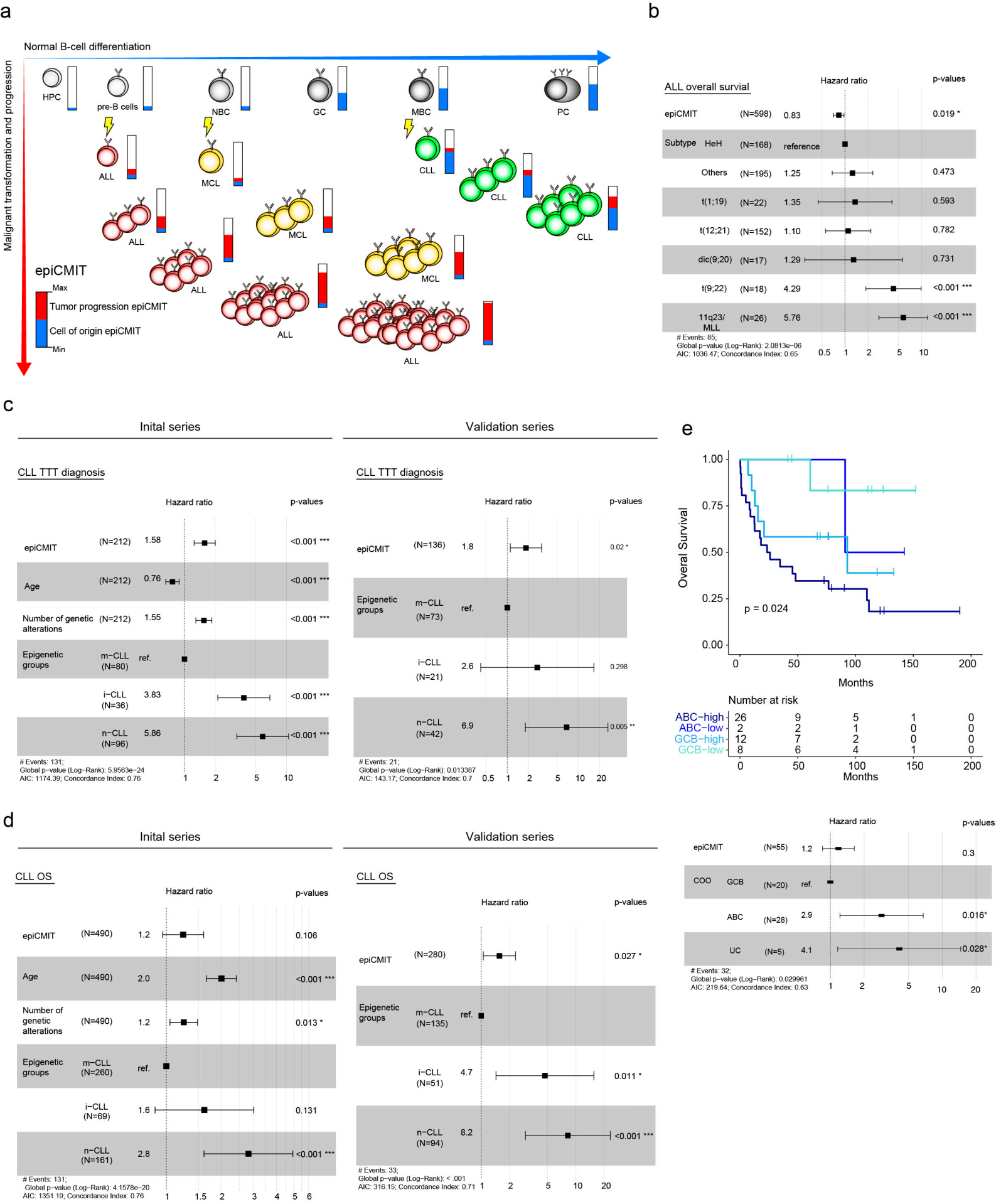
Extended Data Fig. 3



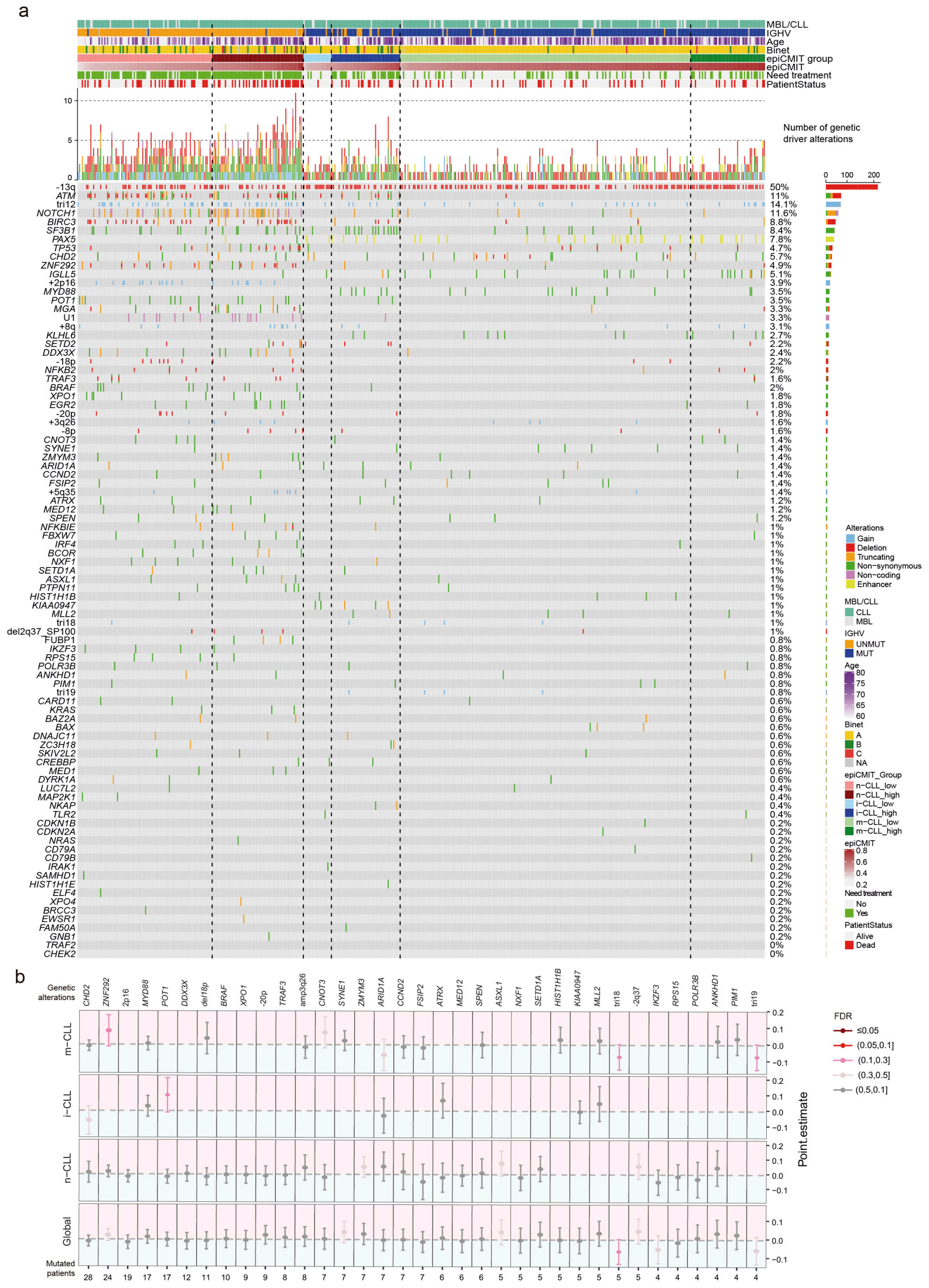
Extended Data Fig. 4



Extended Data Fig. 5



Extended Data Fig. 6



Extended Data Fig. 7