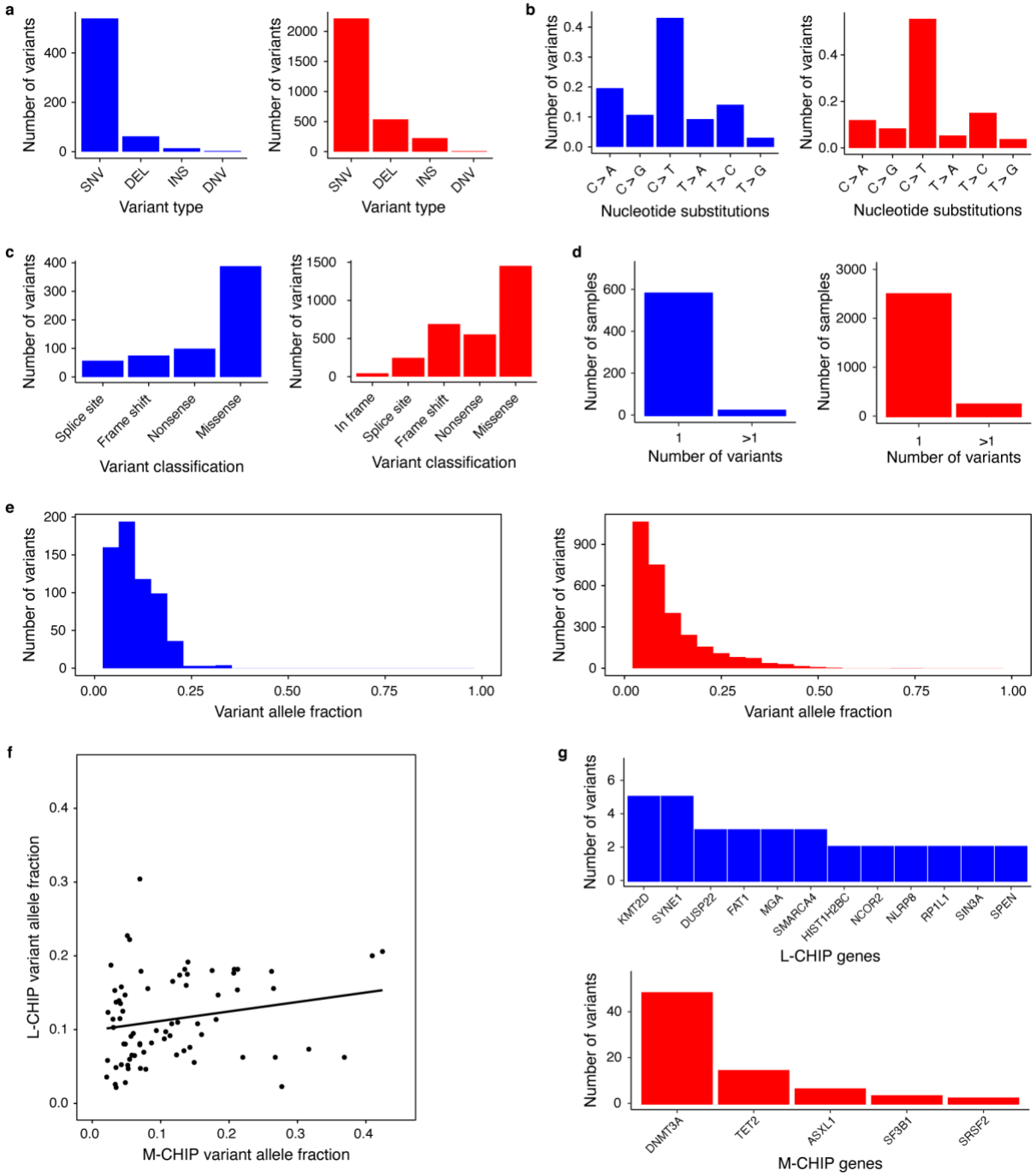
Supplementary Information: Distinction of lymphoid and myeloid clonal hematopoiesis

Abhishek Niroula[1,2,3], Aswin Sekar[2], Mark A. Murakami[2], Mark Trinder[1,4], Mridul Agrawal[2], Waihay J. Wong[1,5], Alexander G. Bick[1,6], Md Mesbah Uddin[1,7], Christopher J. Gibson[1,2], Gabriel K. Griffin[1,5], Michael C. Honigberg[1,8], Seyedeh M. Zekavat [1,9], Kaavya Paruchuri[1,7,8,10], Pradeep Natarajan[1,7,10], Benjamin L. Ebert*[1,2,11]

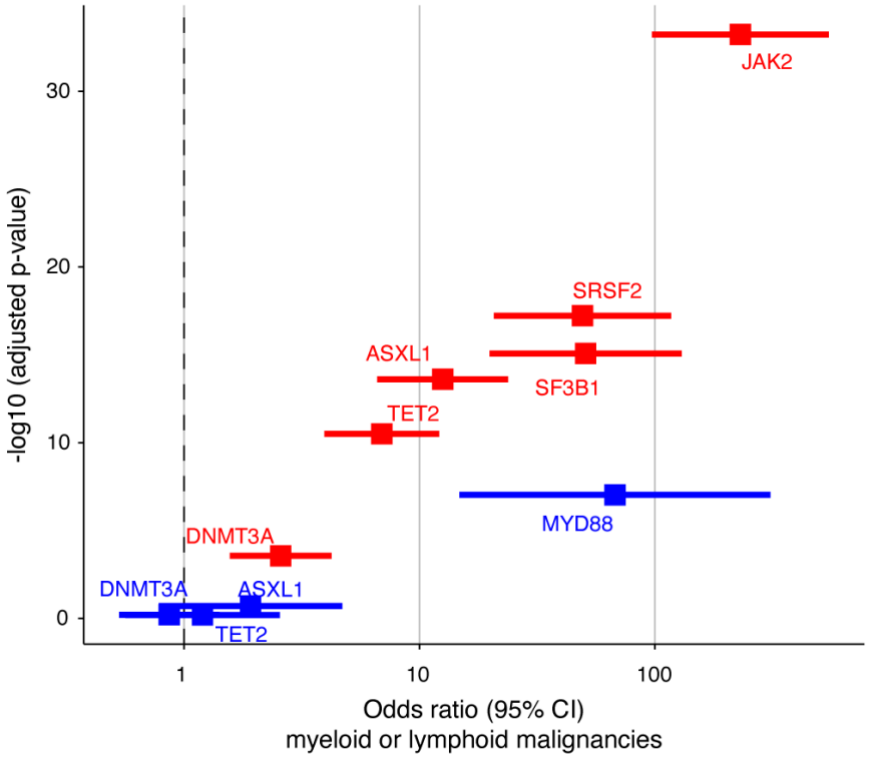
Author affiliations
[1]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[2]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA
[3]Department of Laboratory Medicine, Lund University, Lund, Sweden
[4]Center for Heart Lung Innovation, University of British Columbia, Vancouver
[5]Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA
[6]Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
[7]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA
[8]Cardiology Division, Massachusetts General Hospital, Boston, MA, USA
[9]Yale University School of Medicine, New Haven, CT, USA
[10]Department of Medicine, Harvard Medical School, Boston, MA, USA
[11]Howard Hughes Medical Institute, Bethesda, MD, USA

***Correspondence:**
Benjamin L. Ebert
Dana-Farber Cancer Institute
450 Brookline Ave, D1610A
Boston, MA 02215
Tel: 617-632-1902
Email: benjamin_ebert@dfci.harvard.edu

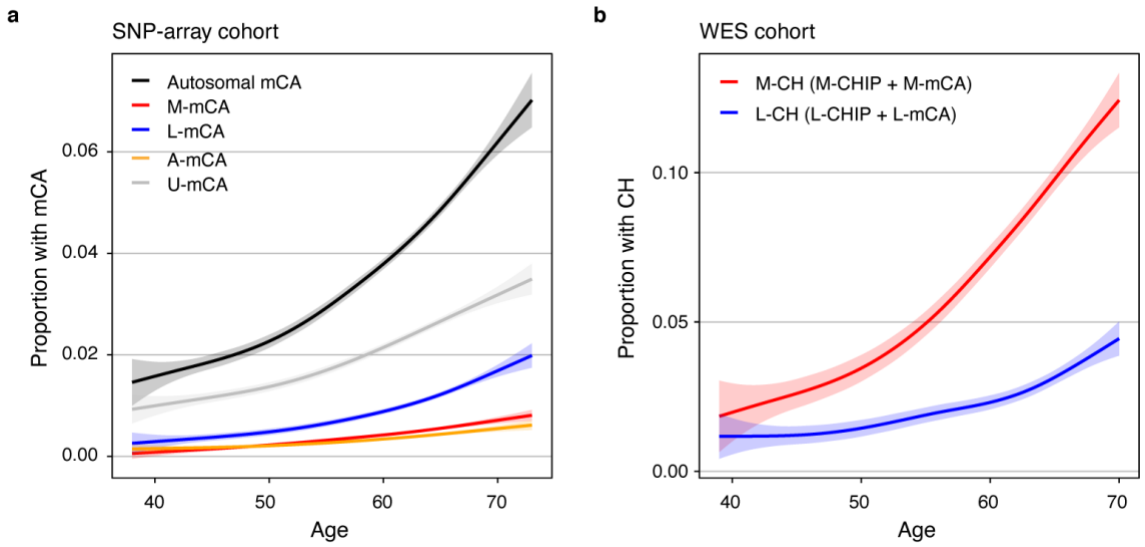
Supplementary Figures



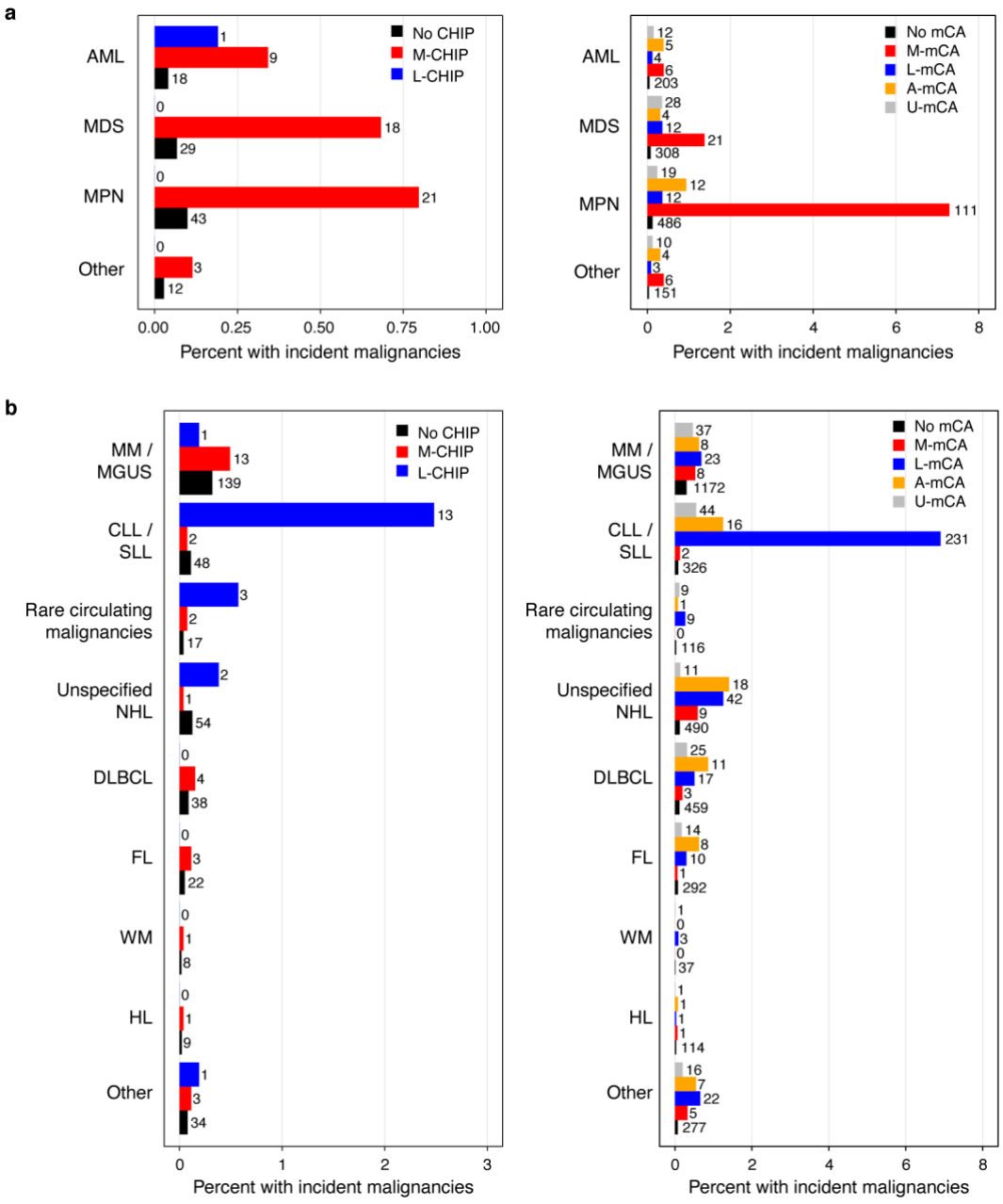
Supplementary Figure S1: Characteristics of somatic variants detected in myeloid and lymphoid driver genes. a) variant type, **b)** nucleotide substitutions for single nucleotide substitutions, **c)** variant classification based on changes in canonical protein sequence, **d)** number of variants detected per individual, **e)** variant allele fraction (VAF) of variants. Characteristics of M-CHIP are shown in red and L-CHIP in blue. **f)** Correlation between clone sizes of M-CHIP and L-CHIP detected in the same individual (n=73). **g)** Myeloid and lymphoid genes mutated among individuals carrying both M-CHIP and L-CHIP.



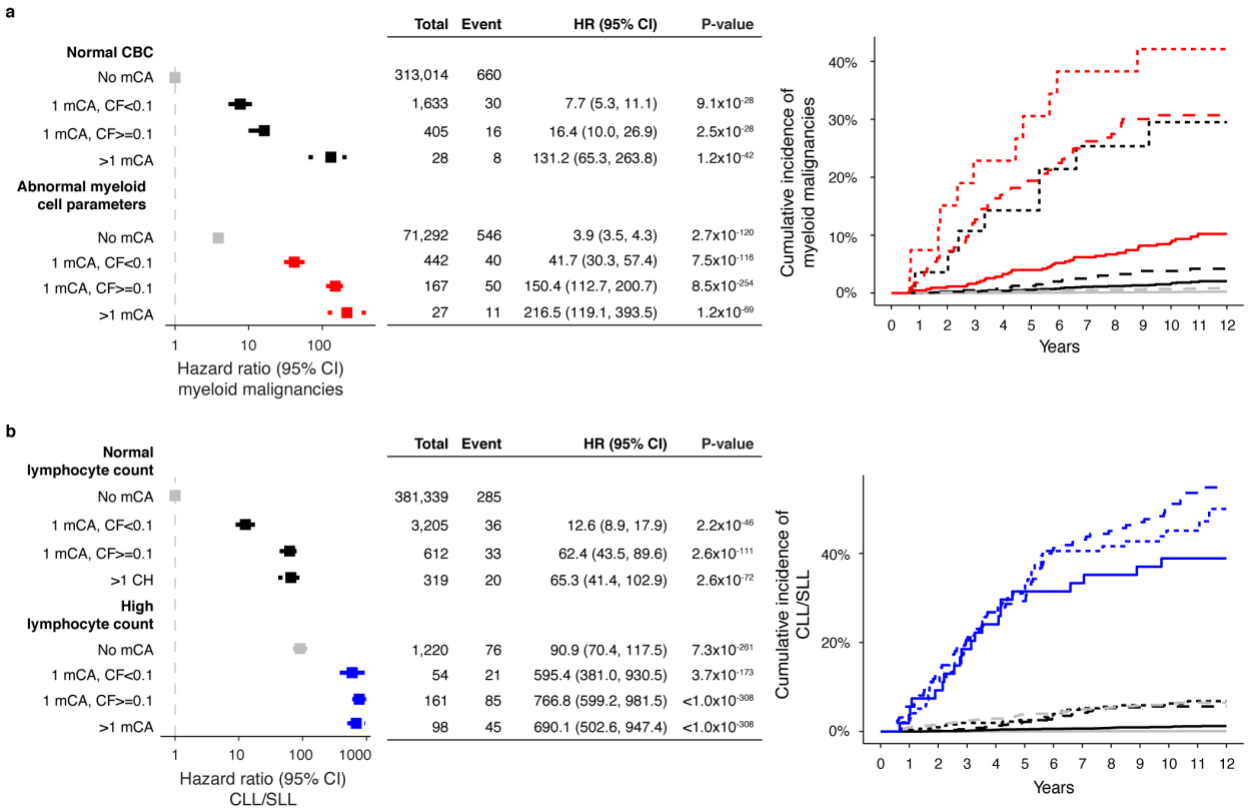
Supplementary Figure S2: Enrichment of CHIP gene mutations in myeloid and lymphoid malignancies. CHIP genes associated with myeloid (red) and lymphoid (blue) malignancies. Only genes mutated in at least 3 individuals who developed an incident myeloid or lymphoid malignancy between six months and 12 years after blood sample collection were analyzed. Tested genes were *DNMT3A* (n=1,827, n myeloid malignancies=18, and n lymphoid malignancies=17), *TET2* (n=510, n myeloid malignancies=14, and n lymphoid malignancies=7), *ASXL1* (n=201, n myeloid malignancies=11, and n lymphoid malignancies=5), *SRSF2* (n=34 and n myeloid malignancies=7), *SF3B1* (n=28 and n myeloid malignancies=6), *JAK2* (n=23, n myeloid malignancies=11), and *MYD88* (n=7, n lymphoid malignancies=3). Data are presented as odds ratio and 95% confidence intervals, computed by generalized linear regression model adjusting for age, age squared, sex, and smoking. CI, confidence interval.



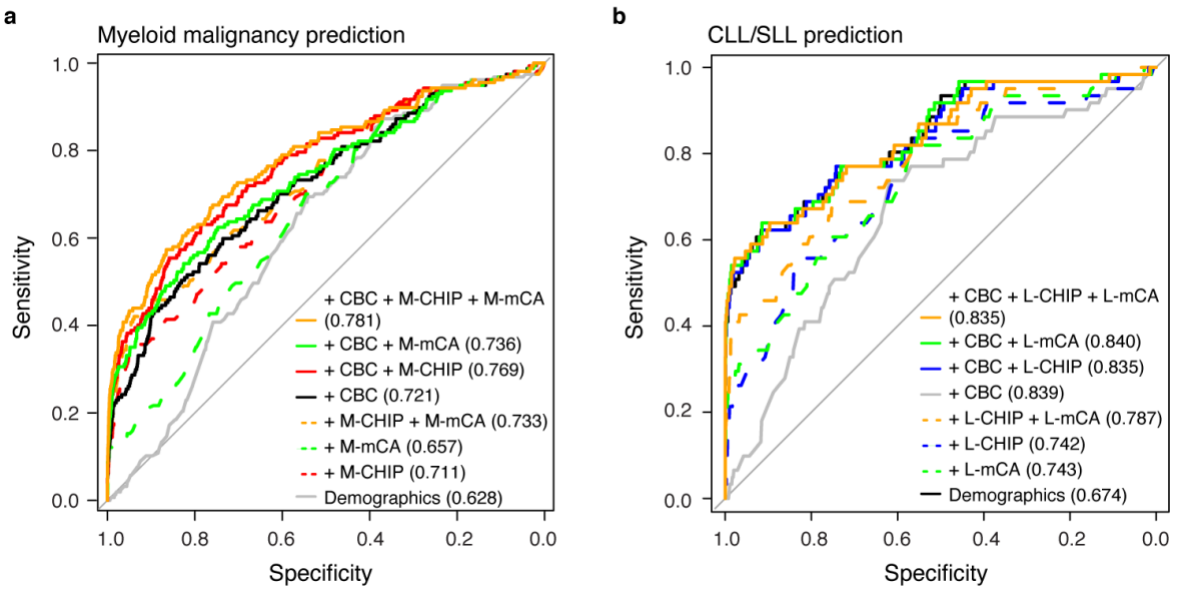
Supplementary Figure S3: Prevalence of CH increases with age. a) All four categories of mCAs are associated with age. The A-mCA (n=1278) is the rarest category of mCAs followed by M-mCA (n=1523), and L-mCA (3345). A large fraction of the mCAs remained unclassified (U-mCA, n=7966). **b)** Frequency of myeloid CH (M-CHIP + M-mCA, n=2975) and lymphoid CH (L-CHIP + L-mCA, n=1045) in the UKB WES cohort. (**a**-**b**) Data is fit with the general additive model using cubic regression splines and the shaded bands indicate the estimated 95% confidence interval.



Supplementary Figure S4: Enrichment of myeloid and lymphoid malignancy sub-types among individuals with CH. a) Frequencies of myeloid malignancy sub-types among CH cases and controls. **b)** Frequencies of lymphoid malignancy sub-types among CH cases and controls. The numbers above the bars indicate the total number of individuals with incident malignancy in those groups. The individuals are categorized based on the first coded hematologic malignancies. MDS, myelodysplastic syndrome; AML, acute myeloid leukemia; MPN, myeloproliferative neoplasms; CLL, chronic lymphocytic leukemia; SLL, small lymphocytic lymphoma; MM, multiple myeloma; MGUS, monoclonal gammopathy of undetermined significance; NHL, non-Hodgkin's lymphoma; DLBCL, diffuse large B-cell lymphoma; FL, follicular lymphoma; WM, Waldenstrom's macroglobulinaemia; HL, Hodgkin's lymphoma; U-mCA, unclassified mCAs.



Supplementary Figure S5: Integration of mCA and blood cell indices stratify risk for developing hematologic malignancies. a) Myeloid CH (M-mCA/A-mCA) and abnormal myeloid cell parameters increase risk of myeloid malignancies. **b)** Lymphoid CH (L-mCA/A-mCA) and elevated lymphocyte count increase risk of CLL/SLL. (**a-b**) Since A-mCA were associated with risk of both myeloid and lymphoid malignancies, these are combined with M-mCA in the analysis of myeloid malignancies and with L-mCA in the analysis of lymphoid malignancies. Data are presented as hazard ratio and 95% confidence intervals, computed by using Cox proportional hazards model adjusting for age, sex, smoking, genetic ethnic ancestry, and genetic principal components 1-5. Red blood cell counts, platelet counts, and neutrophil counts were used to define individuals with abnormal myeloid cell parameters. CBC, complete blood count; HR, hazard ratio; CI, confidence interval; CLL, chronic lymphocytic leukemia; SLL, small lymphocytic lymphoma; VAF, variant allele fraction; CF, cell fraction.



Supplementary Figure S6: **Integration of CH abnormalities and CBC predict the risk of myeloid malignancies and CLL/SLL. a**) Sensitivity and specificity for predicting the incidence of myeloid malignancies computed by 10-fold cross-validation. Individuals with a diagnosis of myeloid malignancy (n=157) and those that did not have a diagnosis of myeloid or lymphoid malignancies (n=44781) were used for training and testing. **b**) Sensitivity and specificity for predicting the incidence of CLL/SLL computed by 10-fold cross-validation. Individuals with a diagnosis of CLL or SLL (n=61) and those that did not have a diagnosis of myeloid or lymphoid malignancies (n=44781) were used for training and testing. (**a-b**) Baseline predictor included demographic characteristics (age, sex, smoking status, and genetic ethnic ancestry). CHIP (number of CHIP mutations and maximum VAF), mCA (number of alterations and maximum cell fraction), and CBC were added as predictors together with demographics. The values inside the parentheses indicate the AUC. Since A-mCA were associated with risk of both myeloid and lymphoid malignancies, these are combined with M-mCA in the analysis of myeloid malignancies and with L-mCA in the analysis of lymphoid malignancies. CLL, chronic lymphocytic leukemia; SLL, small lymphocytic lymphoma; CBC, complete blood count; VAF, variant allele fraction; AUC, area under the curve.

References

58. Morris, S.W., et al. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. Science 263, 1281-1284 (1994).
59. Wellmann, A., et al. Analysis of the t(2;5)(p23;q35) translocation by reverse transcription-polymerase chain reaction in CD30+ anaplastic large-cell lymphomas, in other non-Hodgkin's lymphomas of T-cell phenotype, and in Hodgkin's disease. Blood 86, 2321-2328 (1995).
60. Lamant, L., Dastugue, N., Pulford, K., Delsol, G. & Mariame, B. A new fusion gene TPM3-ALK in anaplastic large cell lymphoma created by a (1;2)(q25;p23) translocation. Blood 93, 3088-3095 (1999).
61. Siebert, R., et al. Complex variant translocation t(1;2) with TPM3-ALK fusion due to cryptic ALK gene rearrangement in anaplastic large-cell lymphoma. Blood 94, 3614-3617 (1999).
62. Chikatsu, N., et al. ALK+, CD30-, CD20- large B-cell lymphoma containing anaplastic lymphoma kinase (ALK) fused to clathrin heavy chain gene (CLTC). Mod Pathol 16, 828-832 (2003).
63. De Paepe, P., et al. ALK activation by the CLTC-ALK fusion is a recurrent event in large B-cell lymphoma. Blood 102, 2638-2641 (2003).
64. Gesk, S., et al. ALK-positive diffuse large B-cell lymphoma with ALK-Clathrin fusion belongs to the spectrum of pediatric lymphomas. Leukemia 19, 1839-1840 (2005).
65. Streubel, B., Vinatzer, U., Willheim, M., Raderer, M. & Chott, A. Novel t(5;9)(q33;q22) fuses ITK to SYK in unspecified peripheral T-cell lymphoma. Leukemia 20, 313-318 (2006).
66. Fujiwara, S.I., et al. High-resolution analysis of chromosome copy number alterations in angioimmunoblastic T-cell lymphoma and peripheral T-cell lymphoma, unspecified, with single nucleotide polymorphism-typing microarrays. Leukemia 22, 1891-1898 (2008).
67. Herens, C., et al. Cyclin D1-negative mantle cell lymphoma with cryptic t(12;14)(p13;q32) and cyclin D2 overexpression. Blood 111, 1745-1746 (2008).
68. Wlodarska, I., et al. Translocations targeting CCND2, CCND3, and MYCN do occur in t(11;14)-negative mantle cell lymphomas. Blood 111, 5683-5690 (2008).
69. Feldman, A.L., et al. Recurrent translocations involving the IRF4 oncogene locus in peripheral T-cell lymphomas. Leukemia 23, 574-580 (2009).
70. Morin, R.D., et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat Genet 42, 181-185 (2010).
71. Rui, L., et al. Cooperative epigenetic modulation by cancer amplicon genes. Cancer Cell 18, 590-605 (2010).
72. Feldman, A.L., et al. Discovery of recurrent t(6;7)(p25.3;q32.3) translocations in ALK-negative anaplastic large cell lymphomas by massively parallel genomic sequencing. Blood 117, 915-919 (2011).
73. Quesada, V., et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet 44, 47-52 (2011).
74. Steidl, C. & Gascoyne, R.D. The molecular pathogenesis of primary mediastinal large B-cell lymphoma. Blood 118, 2659-2669 (2011).
75. Wang, L., et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med 365, 2497-2506 (2011).

76. Greisman, H.A., et al. IgH partner breakpoint sequences provide evidence that AID initiates t(11;14) and t(8;14) chromosomal breaks in mantle cell and Burkitt lymphomas. Blood 120, 2864-2867 (2012).
77. Kridel, R., et al. Whole transcriptome sequencing reveals recurrent NOTCH1 mutations in mantle cell lymphoma. Blood 119, 1963-1971 (2012).
78. Love, C., et al. The genetic landscape of mutations in Burkitt lymphoma. Nat Genet 44, 1321-1325 (2012).
79. Vasmatzis, G., et al. Genome-wide analysis reveals recurrent structural abnormalities of TP63 and other p53-related genes in peripheral T-cell lymphomas. Blood 120, 2280-2289 (2012).
80. Woyach, J.A., Johnson, A.J. & Byrd, J.C. The B-cell receptor signaling pathway as a therapeutic target in CLL. Blood 120, 1175-1184 (2012).
81. Giacomini, C.P., et al. Breakpoint analysis of transcriptional and genomic profiles uncovers novel gene fusions spanning multiple human cancer types. PLoS Genet 9, e1003464 (2013).
82. Ying, C.Y., et al. MEF2B mutations lead to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. Nat Immunol 14, 1084-1092 (2013).
83. Bruno, A., et al. Mutational analysis of primary central nervous system lymphoma. Oncotarget 5, 5065-5075 (2014).
84. Gunawardana, J., et al. Recurrent somatic mutations of PTPN1 in primary mediastinal B cell lymphoma and Hodgkin lymphoma. Nat Genet 46, 329-335 (2014).
85. Hunter, Z.R., et al. The genomic landscape of Waldenstrom macroglobulinemia is characterized by highly recurring MYD88 and WHIM-like CXCR4 mutations, and small somatic deletions associated with B-cell lymphomagenesis. Blood 123, 1637-1646 (2014).
86. Muppidi, J.R., et al. Loss of signalling via Galpha13 in germinal centre B-cell-derived lymphoma. Nature 516, 254-258 (2014).
87. Parrilla Castellar, E.R., et al. ALK-negative anaplastic large cell lymphoma is a genetically heterogeneous disease with widely disparate clinical outcomes. Blood 124, 1473-1480 (2014).
88. Velusamy, T., et al. A novel recurrent NPM1-TYK2 gene fusion in cutaneous CD30-positive lymphoproliferative disorders. Blood 124, 3768-3771 (2014).
89. Braggio, E., et al. Genome-Wide Analysis Uncovers Novel Recurrent Alterations in Primary Central Nervous System Lymphomas. Clin Cancer Res 21, 3986-3994 (2015).
90. Crescenzo, R., et al. Convergent mutations and kinase fusions lead to oncogenic STAT3 activation in anaplastic large cell lymphoma. Cancer Cell 27, 516-532 (2015).
91. Kurtz, D.M., et al. Noninvasive monitoring of diffuse large B-cell lymphoma by immunoglobulin high-throughput sequencing. Blood 125, 3679-3687 (2015).
92. Vater, I., et al. The mutational pattern of primary lymphoma of the central nervous system determined by whole-exome sequencing. Leukemia 29, 677-685 (2015).
93. Yoda, A., et al. Mutations in G protein beta subunits promote transformation and kinase inhibitor resistance. Nat Med 21, 71-75 (2015).
94. Boddicker, R.L., et al. Integrated mate-pair and RNA sequencing identifies novel, targetable gene fusions in peripheral T-cell lymphoma. Blood 128, 1234-1245 (2016).
95. Chapuy, B., et al. Diffuse large B-cell lymphoma patient-derived xenograft models capture the molecular and biological heterogeneity of the disease. Blood 127, 2203-2213 (2016).

96. Chapuy, B., et al. Targetable genetic features of primary testicular and primary central nervous system lymphomas. Blood 127, 869-881 (2016).
97. Fukumura, K., et al. Genomic characterization of primary central nervous system lymphoma. Acta Neuropathol 131, 865-875 (2016).
98. Kridel, R., et al. Histological Transformation and Progression in Follicular Lymphoma: A Clonal Evolution Study. PLoS Med 13, e1002197 (2016).
99. Louissaint, A., Jr., et al. Pediatric-type nodal follicular lymphoma: a biologically distinct lymphoma with frequent MAPK pathway mutations. Blood 128, 1093-1100 (2016).
100. Morin, R.D., et al. Genetic Landscapes of Relapsed and Refractory Diffuse Large B-Cell Lymphomas. Clin Cancer Res 22, 2290-2300 (2016).
101. Rohr, J., et al. Recurrent activating mutations of CD28 in peripheral T-cell lymphomas. Leukemia 30, 1062-1070 (2016).
102. Yoo, H.Y., et al. Frequent CTLA4-CD28 gene fusion in diverse types of T-cell lymphoma. Haematologica 101, 757-763 (2016).
103. Grommes, C., et al. Ibrutinib Unmasks Critical Role of Bruton Tyrosine Kinase in Primary CNS Lymphoma. Cancer Discov 7, 1018-1029 (2017).
104. Lazarian, G., Guieze, R. & Wu, C.J. Clinical Implications of Novel Genomic Discoveries in Chronic Lymphocytic Leukemia. J Clin Oncol 35, 984-993 (2017).
105. Pedersen, M.B., et al. DUSP22 and TP63 rearrangements predict outcome of ALK-negative anaplastic large cell lymphoma: a Danish cohort study. Blood 130, 554-557 (2017).
106. Arthur, S.E., et al. Genome-wide discovery of somatic regulatory variants in diffuse large B-cell lymphoma. Nat Commun 9, 4001 (2018).
107. Hunter, Z.R., et al. Insights into the genomic landscape of MYD88 wild-type Waldenstrom macroglobulinemia. Blood Adv 2, 2937-2946 (2018).
108. Krumbholz, M., et al. Characterization and diagnostic application of genomic NPM-ALK fusion sequences in anaplastic large-cell lymphoma. Oncotarget 9, 26543-26555 (2018).
109. Vallois, D., et al. RNA fusions involving CD28 are rare in peripheral T-cell lymphomas and concentrate mainly in those derived from follicular helper T cells. Haematologica 103, e360-e363 (2018).
110. Chapuy, B., et al. Genomic analyses of PMBL reveal new drivers and mechanisms of sensitivity to PD-1 blockade. Blood 134, 2369-2382 (2019).
111. de Araujo, E.D., et al. Structural and functional consequences of the STAT5B(N642H) driver mutation. Nat Commun 10, 2517 (2019).
112. Gimenez, N., et al. Mutations in the RAS-BRAF-MAPK-ERK pathway define a specific subgroup of patients with adverse clinical features and provide new therapeutic options in chronic lymphocytic leukemia. Haematologica 104, 576-586 (2019).
113. Gruber, M., et al. Growth dynamics in naturally progressing chronic lymphocytic leukaemia. Nature 570, 474-479 (2019).
114. Guieze, R., et al. Mitochondrial Reprogramming Underlies Resistance to BCL-2 Inhibition in Lymphoid Malignancies. Cancer Cell 36, 369-384 e313 (2019).
115. Liang, W.S., et al. Comprehensive Genomic Profiling of Hodgkin Lymphoma Reveals Recurrently Mutated Genes and Increased Mutation Burden. Oncologist 24, 219-228 (2019).
116. Liu, Y. & Barta, S.K. Diffuse large B-cell lymphoma: 2019 update on diagnosis, risk stratification, and treatment. Am J Hematol 94, 604-616 (2019).

117. Martin-Garcia, D., et al. CCND2 and CCND3 hijack immunoglobulin light-chain enhancers in cyclin D1(-) mantle cell lymphoma. Blood 133, 940-951 (2019).
118. Onaindia, A., et al. DUSP22-rearranged anaplastic lymphomas are characterized by specific morphological features and a lack of cytotoxic and JAK/STAT surrogate markers. Haematologica 104, e158-e162 (2019).
119. Wright, G.W., et al. A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications. Cancer Cell 37, 551-568 e514 (2020).