Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence

Giulio Genovese, Ph.D., **Anna K. Kähler, Ph.D.**, **Robert E. Handsaker, B.S.**, **Johan Lindberg, Ph.D.**, **Samuel A. Rose, B.S.**, **Samuel F. Bakhoum, M.D., Ph.D.**, **Kimberly Chambert, M.S.**, **Eran Mick, B.S.**, **Benjamin M. Neale, Ph.D.**, **Menachem Fromer, Ph.D.**, **Shaun M. Purcell, Ph.D.**, **Oscar Svantesson, M.S.**, **Mikael Landén, Ph.D.**, **Martin Höglund, M.D., Ph.D.**, **Sören Lehmann, M.D., Ph.D.**, **Stacey B. Gabriel, Ph.D.**, **Jennifer L. Moran, Ph.D.**, **Eric S. Lander, Ph.D.**, **Patrick F. Sullivan, M.D.**, **Pamela Sklar, M.D., Ph.D.**, **Henrik Grönberg, M.D., Ph.D.**, **Christina M. Hultman, Ph.D.**, and **Steven A. McCarroll, Ph.D.**

Abstract

BACKGROUND—Cancers arise from multiple acquired mutations, which presumably occur over many years. Early stages in cancer development might be present years before cancers become clinically apparent.

METHODS—We analyzed data from whole-exome sequencing of DNA in peripheral-blood cells from 12,380 persons, unselected for cancer or hematologic phenotypes. We identified somatic mutations on the basis of unusual allelic fractions. We used data from Swedish national patient registers to follow health outcomes for 2 to 7 years after DNA sampling.

RESULTS—Clonal hematopoiesis with somatic mutations was observed in 10% of persons older than 65 years of age but in only 1% of those younger than 50 years of age. Detectable clonal expansions most frequently involved somatic mutations in three genes (*DNMT3A, ASXL1*, and *TET2*) that have previously been implicated in hematologic cancers. Clonal hematopoiesis was a strong risk factor for subsequent hematologic cancer (hazard ratio, 12.9; 95% confidence interval, 5.8 to 28.7). Approximately 42% of hematologic cancers in this cohort arose in persons who had clonality at the time of DNA sampling, more than 6 months before a first diagnosis of cancer. Analysis of bone marrow–biopsy specimens obtained from two patients at the time of diagnosis of acute myeloid leukemia revealed that their cancers arose from the earlier clones.

CONCLUSIONS—Clonal hematopoiesis with somatic mutations is readily detected by means of DNA sequencing, is increasingly common as people age, and is associated with increased risks of hematologic cancer and death. A subset of the genes that are mutated in patients with myeloid cancers is frequently mutated in apparently healthy persons; these mutations may represent

characteristic early events in the development of hematologic cancers. (Funded by the National Human Genome Research Institute and others.)

T<small>HE DEVELOPMENT OF DISEASE OFTEN</small> involves dynamic processes that begin years or decades before the clinical onset. In many cases, however, the process of pathogenesis goes undetected until after the patient has symptoms and presents with clinically apparent disease.

Cancer arises owing to the combined effects of multiple somatic mutations, which are likely to be acquired at different times.[1] Early mutations may be present many years before disease develops. In some models of cancer development, early mutations lead to clonal expansions by stem cells or other progenitor cells.[2] Such clonal expansions greatly increase the likelihood that later, cooperating mutations would arise in cells that already contain the earlier, initiating mutations. To understand the pathogenesis of proliferative diseases, it is important to know the extent to which clonal expansions occur and precede cancer.

Hematopoietic stem-cell population dynamics may precede many hematologic cancers, including myeloproliferative neoplasms,[3] myelodysplastic syndromes,[4] acute myeloid leukemia (AML),[5,6] and chronic lymphocytic leukemia.[7] For example, in some patients, stem cells carrying a subset of the mutations present in the cancer cells are able to survive chemotherapy; subsequently, these cells acquire novel mutations, triggering a relapse.[8-10] This suggests that a clonally expanded stem-cell population may have existed before the cancer developed.

Clonal mosaicism for large chromosomal abnormalities, reflecting expansion of a specific cellular clone, appears to arise in approximately 2% of elderly persons and is a risk factor for later hematopoietic cancers.[11-15] In principle, clonal expansion among hematopoietic stem cells — a phenomenon termed "clonal hematopoiesis" — could be much more common[16] and only occasionally accompanied by chromosomal abnormalities.

Many studies today sequence blood-derived DNA from thousands of persons to identify inherited risk factors for common diseases. We reasoned that such data offered the opportunity to test the hypothesis that clonal expansions with somatic mutations are common and often precede blood cancers and to identify the genes in which mutations drive clonal expansions.

METHODS

STUDY PARTICIPANTS AND WHOLE-EXOME SEQUENCING OF BLOOD-DERIVED DNA

We obtained blood samples for DNA sequencing from 12,380 Swedish persons (mean age at time of sample collection, 55 years [range, 19 to 93]) (see Fig. S1 in the Supplementary Appendix, available with the full text of this article at NEJM.org). This cohort included 6245 controls, 4970 persons with schizophrenia,[17,18] and 1165 persons with bipolar disorder (Table S1 in the Supplementary Appendix). We treated case patients and controls as a single cohort for all analyses presented below, because none of the mutational variables analyzed showed any significant relationship to psychiatric diagnosis after we controlled for other factors such as age and smoking status. DNA was extracted directly from peripheral venous blood samples. All procedures were approved by the relevant ethics committees, and written

informed consent was obtained from all participants. Methods of DNA sequencing and analysis are described in detail in the Supplementary Appendix.

IDENTIFICATION OF SOMATIC MUTATIONS

We developed a strategy for identifying somatic mutations on the basis of allelic fractions. Assuming that a somatic mutation would be present in only a subset of the cells from which we obtained DNA for analysis, we predicted that the mutant allele would be present in less than 50% of the sequencing reads arising from that genomic site (Fig. 1A and 1B). We describe this analytic approach in detail in the Supplementary Appendix.

ANALYSES OF SUBSEQUENT MEDICAL OUTCOMES AND CANCERS

Medical histories (from 1965 to 2012) for 11,164 participants were extracted from the Swedish national inpatient, outpatient, and cause-of-death registers. We used these data to follow health outcomes for 2 to 7 years after DNA sampling. For 2 participants who received a diagnosis of cancer 2 and 34 months after the initial DNA sampling, bone marrow–biopsy specimens (obtained at the time of diagnosis) were analyzed by means of whole-exome and whole-genome sequencing.

RESULTS

CHARACTERISTIC MUTATIONS AND CANDIDATE DRIVERS IN CLONAL HEMATOPOIESIS

We analyzed whole-exome sequencing data from 12,380 persons and identified 3111 putative somatic mutations on the basis of their presence at unusual allelic fractions, for a frequency of approximately 1 putative somatic mutation for every 4 participants. For all 65 mutations tested, molecular validation confirmed that the mutant allele was present at a low allelic fraction (<50%) and thus could not have been inherited (Fig. S7 in the Supplementary Appendix).

The vast majority of the mutations were dispersed across the genome. However, four genes (*DNMT3A, TET2, ASXL1*, and *PPM1D*) had disproportionately high numbers of somatic mutations. Although more than 94% of the mutations observed across the genome were missense and synonymous changes, the somatic mutations observed in *DNMT3A, TET2, ASXL1*, and *PPM1D* had a strong tendency to disrupt gene protein-coding sequences by introducing a frame-shift, nonsense, or splice-site disruption (Fig. 2A). Three of these four genes — *DNMT3A, TET2*, and *ASXL1* — also tend to harbor such mutations in hematologic cancers[19-21] and are proposed to function as epigenetic regulators.[22]

Mutations in *PPM1D*, which functions as a regulator of tumor-suppressor protein p53,[23] have been described more frequently in nonhematologic cancers. Of the 15 protein-truncating mutations observed in *PPM1D*, 12 occurred in the last exon, which is also the site of protein-truncating mutations described in patients with cancer.[24-27] Loss of the C-terminal localization domain of *PPM1D* is reported to activate *PPM1D*, repress p53, and thereby impair the p53-dependent G1 checkpoint, promoting proliferation.[26]

Somatic mutations in *DNMT3A* were also characterized by a significant excess of missense mutations (P<0.001) (Fig. 2A), all localized in exons 7 to 23, and were enriched for

cysteine-forming mutations (Fig. S9 in the Supplementary Appendix). Such mutations potentially exert a dominant-negative effect on the tetrameric *DNMT3A* protein complex (see the Supplementary Appendix for details).

Because *DNMT3A, TET2*, and *ASXL1* are frequently mutated in hematologic cancers,[28,29] we hypothesized that other recurring cancer mutations might also promote clonal hematopoiesis. We therefore considered 208 specific variants reported as recurring (found in ≥7 patients) in hematopoietic and lymphoid cancers, as listed in the Catalogue of Somatic Mutations in Cancer[30] (see the Supplementary Appendix for details). We found 98 instances of these recurring mutations, including *JAK2* mutation p.V617F[31] in 24 participants, *DNMT3A* mutation p.R882H[32] in 15 participants, and *SF3B1* mutation p.K700E[33] in 9 participants.

Disruptive mutations in *DNMT3A, TET2, ASXL1*, and *PPM1D*, missense mutations in *DNMT3A*, and the recurring cancer-associated mutations comprised a set of 327 candidate driver somatic mutations for clonal hematopoiesis in 308 participants (Fig. 2B, and Table S3 in the Supplementary Appendix), with 18 participants carrying multiple candidate drivers (Fig. 2C). *DNMT3A* had the most observed mutations (190), followed by *ASXL1* (35) and *TET2* (31). Mutations in *TET2* and *ASXL1* were probably undercounted for technical reasons (Fig. S6B and S6C in the Supplementary Appendix).

CLONAL HEMATOPOIESIS WITH UNKNOWN DRIVERS

Somatic mutations may be either drivers, which contribute to clonal expansion, or passengers, which do not; either could indicate the presence of a clone. Participants having clonal hematopoiesis with candidate drivers tended to carry more additional putative somatic mutations than participants without candidate drivers (mean number of additional mutations, 1.5 vs. 0.2; P<0.001 after correction for age) (Fig. S11A in the Supplementary Appendix).

A total of 459 participants had multiple putative somatic mutations without any of the candidate drivers described above. Estimates of allelic fraction were more similar in pairs of somatic mutations observed in the same person than in pairs of somatic mutations observed in different persons (P<0.001 by the Mann–Whitney test), a finding consistent with the possibility that the mutations were present in the same clone.

To consider cases of clonal hematopoiesis without obvious driver mutations, we sought to define a highly specific criterion for clonal hematopoiesis that depended on the number of mutations alone, rather than on the identity of the mutations. Of 11,845 participants with sequencing data of sufficient quality for the detection of somatic mutations, 9927 had no putative somatic mutations, 1333 had 1 mutation, 313 had 2 mutations, and 272 had 3 to 18 mutations. This distribution suggested that a random (Poisson) process (with a constant mean) could not explain the surprisingly high numbers of participants with 3 to 18 detectable mutations. In our analyses below, we classified 195 participants as having clonal hematopoiesis with unknown drivers. In some cases of clonal hematopoiesis with unknown drivers, additional analysis suggested potential candidate drivers (Fig. S12 in the Supplementary Appendix).

CLONAL HEMATOPOIESIS AND ADVANCING AGE

Detectable clonal hematopoiesis with candidate drivers was rare among younger participants (occurring in 0.7% of participants younger than 50 years of age) but much more common among older participants (occurring in 5.7% of participants older than 65 years of age) (Fig. 2D). Reflecting this relationship, participants having clonal hematopoiesis with candidate drivers were, on average, older than those without detectable putative somatic mutations (mean age, 64 years vs. 55 years; P<0.001); each of the most common driver genes (*DNMT3A, ASXL1, TET2, PPMD1*, and *JAK2*) also tended to have detectable somatic mutations in older persons (Fig. S11B in the Supplementary Appendix).

Given that 459 participants had multiple somatic mutations without clear candidate driver mutations, we sought to understand the extent to which clonal hematopoiesis with unknown drivers arises dynamically over the life span (as opposed to being a lifelong property — e.g., due to somatic mutations that occurred in embryonic development[34]). The observation of a single somatic mutation in the exome was common at all ages, in contrast to the strongly age-dependent acquisition of clonal hematopoiesis with candidate drivers (Fig. S13 in the Supplementary Appendix). However, the presence of two detectable putative somatic mutations was more age-dependent, occurring in 1.3% of persons younger than 50 years of age as compared with 4.0% of those older than 65 years of age (Fig. S13 in the Supplementary Appendix). The presence of three or more putative somatic mutations (our criterion for clonal hematopoiesis with unknown drivers) was more strongly age-dependent, occurring in only 0.3% of persons younger than 50 years of age but in 4.6% of those older than 65 years of age (Fig. 2D). This age trajectory is similar to that of clonal hematopoiesis with candidate drivers.

Overall, clonal hematopoiesis with candidate or unknown drivers was observed in 0.9% of participants younger than 50 years of age but in 10.4% of those older than 65 years of age (Fig. 2D). For participants with clonal hematopoiesis, the average number of detected putative somatic mutations also increased with age (P<0.001 by linear regression) (Fig. S14 in the Supplementary Appendix).

CLONAL HEMATOPOIESIS AND SUBSEQUENT HEMATOLOGIC CANCER AND DEATH

Participants with clonal hematopoiesis were substantially more likely to receive a first diagnosis of hematologic cancer 6 months or more after DNA sampling (Table S6 in the Supplementary Appendix) than participants without any detectable putative somatic mutations (hazard ratio, 12.9; 95% confidence interval [CI], 5.8 to 28.7; P<0.001 by Cox proportional-hazards analysis of time to a diagnosis of hematologic cancer, with adjustment for age and sex) (Fig. 3A). Participants with candidate drivers and those with unknown drivers had similarly elevated risks (Fig. 3B). Of the 31 participants who received a diagnosis of hematologic cancer more than 6 months after DNA sampling, 13 participants (42%) had shown clonal hematopoiesis in their initial DNA sample (Fig. 3C). Participants with clonal hematopoiesis were also more likely than those without clonal hematopoiesis to have a history of hematologic cancer, though this did not contribute to the above result (Table S7 in the Supplementary Appendix).

Participants with clonal hematopoiesis had reduced overall survival (hazard ratio for death, 1.4; 95% CI, 1.0 to 1.8; P = 0.03 by Cox proportional-hazards model, with adjustment for age and sex) (Fig. 3D and 3E). A total of 54 participants with clonal hematopoiesis died during follow-up (Table S8 in the Supplementary Appendix). In our cohort, the reduced overall survival was explained by deaths from cancer and by an association of clonal hematopoiesis with smoking (odds ratio, 2.2; 95% CI, 1.4 to 3.4; P<0.001) (see the Supplementary Appendix for details).

MALIGNANT CLONES IN DNA SAMPLES

Two participants with clonal hematopoiesis received a diagnosis of myeloid cancer just 2 months after DNA sampling. We hypothesized that the clone inferred from whole-exome sequencing analysis might have been the malignant clone at a preclinical stage (see the Supplementary Appendix for details).

Whole-genome sequencing analysis of the preclinical blood DNA sample revealed 1153 putative somatic mutations at a characteristic frequency in Participant 1 and 660 such mutations in Participant 2 (Fig. 4A and 4B), evidence that a clone had amplified from a single cell and that was consistent with the number of mutations observed in AML genomes.[29] The whole-genome sequencing data showed that many known pathogenic mutations were present at the characteristic allele frequency of the clone (Fig. 4A and 4B). At 2 months before receiving a diagnosis of the myelodysplastic syndrome, Participant 1 carried mutations in *RUNX1, STAG2, SRSF2*, and *TET2* at the allelic fractions characteristic of the clone and also carried an *ASXL1* mutation at a slightly higher allelic fraction (potentially consistent with a founder mutation). Mutations in *ASXL1, RUNX1*, and *STAG2* tend to co-occur in myelodysplastic syndromes.[35] At 2 months before receiving a diagnosis of AML, Participant 2 carried 2 mutations in *CEBPA*: an in-frame C-terminal 33-bp insertion and a frame-shift N-terminal deletion. These 2 types of *CEBPA* mutation frequently co-occur in AML.[36]

GENETIC RELATIONSHIP OF LEUKEMIA TO EARLIER CLONES

We analyzed bone marrow–biopsy specimens obtained at the time of diagnosis from two participants: Participant 2, who carried the two *CEBPA* mutations, and Participant 3, who received a diagnosis of AML 34 months after DNA sampling (see the Supplementary Appendix for details).

Analysis of the specimen obtained from Participant 2 confirmed the presence of the mutations detected in the earlier DNA sample, including the two *CEBPA* mutations and three passenger mutations now observed at higher allelic fractions (20.5% vs. 15.5% 2 months earlier; estimate based on the three single-nucleotide substitutions). Cancers characterized by pairs of *CEBPA* mutations tend to have a favorable prognosis,[37] and indeed this patient had a complete remission after chemotherapy and did not have a relapse.

The hematologic cancer in Participant 3 appeared to be a subclone of the clone detected initially, which had undergone additional clonal evolution in the intervening 34 months. A *TP53* p.R248Q mutation, present at an allelic fraction of 24% in the initial sample, had

expanded to 86%, which was consistent with the loss of the other copy of chromosome 17 and with the 86% blast count in the bone marrow–biopsy specimen. DNA analysis of the biopsy specimen indicated losses of chromosome 17 and 5q (results that were consistent with the karyotype findings) and a complex pattern of gains and losses on chromosomes 12, 13, 16, and 19 (Fig. S15 in the Supplementary Appendix). Losses of 17 and 5q are common in persons with *TP53* mutations.[38] By using these segmental losses to distinguish between alleles on the lost segments and alleles on the retained segments, we estimated that there were already losses of 5q and 17 at low allelic fractions (8% and 3% of cells, respectively) in the initial DNA sample but with no losses on chromosomes 12, 13, 16, and 19 or with losses at an undetectable frequency at that time (Fig. S16 in the Supplementary Appendix). Because the biopsy specimen showed all these events at high allelic fractions, we concluded that these mutations arose in a series of subclones (Fig. 4C). The patient died 2 months after diagnosis.

DISCUSSION

We observed clonal hematopoiesis with somatic mutations in 10% of the elderly study participants, with increasing frequency with age (Fig. 2D). Most such clonal expansions appeared to involve driver genes and mutations that are also driver mutations in hematologic cancers (Fig. 2A and 2B). We found the presence of such clones to be a risk factor for subsequent hematologic cancers (hazard ratio, 12.9; 95% CI, 5.8 to 28.7) (Fig. 3A and 3B) and death (hazard ratio, 1.4; 95% CI, 1.0 to 1.8) (Fig. 3D and 3E).

Most cases of clonal hematopoiesis appeared to involve mutations in a specific subset of the genes recognized as drivers of blood cancers,[22] such as *DNMT3A, ASXL1*, and *TET2* (Fig. 2A). Other common mutational drivers of such cancers — for example, activating mutations in *FLT3* and *NPM1*[29] — were not observed in these subclinical clonal expansions. Such data suggest that mutations in *DNMT3A, ASXL1*, and *TET2* are often initiating mutations that remain in subclinical states for long periods; *FLT3* and *NPM1* mutations may tend to be later, cooperating events. Such an inference would align with data from studies involving patients with cancer and mouse models.[12-14] Functional experiments have shown that loss of *DNMT3A* impairs differentiation of hematopoietic stem cells, resulting in an increase in the number of such cells in the bone marrow,[39] and that loss of *TET2* results in increased self-renewal of hematopoietic stem cells and a competitive growth advantage.[40]

We detected clonal hematopoiesis in 42% of the participants who received a diagnosis of cancer more than 6 months later (Fig. 3C), and such clones were a strong risk factor for these cancers. Our results raise the question of whether DNA sequencing of blood samples on a regular basis could enable early detection of blood cancers. On the basis of the current data, we believe that such an approach would be premature. Clonal hematopoiesis with somatic mutations was detected in 10% of elderly participants, and the absolute risk of conversion from clonal hematopoiesis to hematologic cancer was modest (1.0% per year). Moreover, there are no currently available interventions that seem suitable for large cohorts of patients with only a modest likelihood of cancer. Caution is also warranted when cancer-associated mutations are observed as an incidental finding in other studies or diagnostic

tests: our results suggest that such findings may be common and do not justify a diagnosis of hematologic cancer, though they do signify an elevated risk.

In the future, however, it may be possible to refine our DNA analysis in order to develop strategies for early detection and even prevention of hematologic cancer. DNA analysis will offer three important capabilities: the ability to ascertain high-risk states, the ability to monitor the progression or remission of these states, and the ability to ascertain follow-on, transforming mutations before clinically apparent illness. Several important research directions could bring DNA sequencing for clonal hematopoiesis closer to clinical usefulness. First, some somatic mutations are likely to be associated with a particularly high risk of subsequent cancer; larger studies could identify such mutations. Second, single-cell analyses might identify high-risk combinations of mutations occurring in the same cells. Third, the sequencing of specific cell types might identify mutation–cell-type combinations with increased predictive value. Fourth, initial detection of clonal hematopoiesis might justify periodic screening for the presence of cooperating mutations at low allele frequencies that could presage cancer (Fig. 4C). Larger studies will be needed to gather the necessary data to evaluate these possibilities.

In addition, the use of DNA sequencing to identify at-risk cohorts and monitor clonal expansions, as reported here, could facilitate clinical trials of strategies to reduce the risk of progression to cancer. In other areas of medicine, such as cardiovascular disease, biomarkers (e.g., low-density lipoprotein cholesterol) that identify high-risk groups of patients and are used to obtain rapid information on the efficacy of interventions have been critical in developing preventive therapies. It will also be important to understand the extent to which clonal hematopoiesis is a marker for the declining health of hematopoietic stem-cell populations, potentially reflecting aging, attrition, and a declining ability to contain new neoplasms.

These findings illustrate a way in which DNA sequencing offers information about dynamic processes that progress over the course of a person's life and are predictive of clinical disease and death. We speculate that an improved ability to predict future harm in asymptomatic persons may eventually replace the traditional dichotomy between illness and health with a continuum of ascertainable genomic states that are associated with elevated risks of future illness.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Appendix

The authors' affiliations are as follows: the Stanley Center for Psychiatric Research (G.G., R.E.H., S.A.R., K.C., B.M.N., J.L.M., S.A.M.) and Program in Medical and Population Genetics (G.G., R.E.H., S.A.M.), Broad Institute of the Massachusetts Institute of Technology and Harvard (S.B.G., E.S.L.), Cambridge, MA; the Departments of Genetics (G.G., R.E.H., S.A.M.) and Systems Biology (E.M.), Harvard Medical School, and the Analytic and Translational Genetics Unit and the Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital (B.M.N.) — both in Boston; the Department of Medical Epidemiology and Biostatistics (A.K.K., J.L., O.S., P.F.S., H.G., C.M.H.) and Science for Life Laboratory (J.L.), Karolinska Institutet, and the Hematology Center and Center for Hematology and Regenerative Medicine, Karolinska University Hospital (S.L.), Stockholm, the Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg (M.L.), and the Department of Medical Sciences and Division of Hematology, University Hospital, Uppsala (M.H.) — all in Sweden; the Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center (S.F.B.), and the Division of Psychiatric Genomics, Department of Psychiatry, and Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai (M.F., S.M.P., P.S.) — both in New York; and the Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill (P.F.S.).

References

1. Nowell PC. The clonal evolution of tumor cell populations. Science. 1976; 194:23–8. [PubMed: 959840]

2. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature. 2001; 414:105–11. [PubMed: 11689955]

3. Jamieson CHM, Gotlib J, Durocher JA, et al. The JAK2 V617F mutation occurs in hematopoietic stem cells in polycythemia vera and predisposes toward erythroid differentiation. Proc Natl Acad Sci U S A. 2006; 103:6224–9. [PubMed: 16603627]

4. Jaiswal S, Ebert BL. MDS is a stem cell disorder after all. Cancer Cell. 2014; 25:713–4. [PubMed: 24937455]

5. Potter NE, Greaves M. Cancer: persistence of leukaemic ancestors. Nature. 2014; 506:300–1. [PubMed: 24522525]

6. Vasanthakumar A, Godley LA. On the origin of leukemic species. Cell Stem Cell. 2014; 14:421–2. [PubMed: 24702991]

7. Damm F, Mylonas E, Cosson A, et al. Acquired initiating mutations in early hematopoietic cells of CLL patients. Cancer Discov. 2014; 9:1088–101. [PubMed: 24920063]

8. Ding L, Ley TJ, Larson DE, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012; 481:506–10. [PubMed: 22237025]

9. Shlush LI, Zandi S, Mitchell A, et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature. 2014; 506:328–33. [PubMed: 24522528]

10. Corces-Zimmerman MR, Hong W-J, Weissman IL, Medeiros BC, Majeti R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. Proc Natl Acad Sci U S A. 2014; 111:2548–53. [PubMed: 24550281]

11. Forsberg LA, Rasi C, Razzaghian HR, et al. Age-related somatic structural changes in the nuclear genome of human blood cells. Am J Hum Genet. 2012; 90:217–28. [PubMed: 22305530]

12. Laurie CC, Laurie CA, Rice K, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat Genet. 2012; 44:642–50. [PubMed: 22561516]

13. Jacobs KB, Yeager M, Zhou W, et al. Detectable clonal mosaicism and its relationship to aging and cancer. Nat Genet. 2012; 44:651–8. [PubMed: 22561519]

14. Schick UM, McDavid A, Crane PK, et al. Confirmation of the reported association of clonal chromosomal mosaicism with an increased risk of incident hematologic cancer. PLoS One. 2013; 8(3):e59823. [PubMed: 23533652]

15. Forsberg LA, Rasi C, Malmqvist N, et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. Nat Genet. 2014; 46:624–8. [PubMed: 24777449]

16. Busque L, Mio R, Mattioli J, et al. Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. Blood. 1996; 88:59–65. [PubMed: 8704202]

17. Ripke S, O'Dushlaine C, Chambert K, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet. 2013; 45:1150–9. [PubMed: 23974872]

18. Purcell SM, Moran JL, Fromer M, et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature. 2014; 506:185–90. [PubMed: 24463508]

19. Ley TJ, Ding L, Walter MJ, et al. *DNMT3A* mutations in acute myeloid leukemia. N Engl J Med. 2010; 363:2424–33. [PubMed: 21067377]

20. Delhommeau F, Dupont S, Della Valle V, et al. Mutation in *TET2* in myeloid cancers. N Engl J Med. 2009; 360:2289–301. [PubMed: 19474426]

21. Gelsi-Boyer V, Trouplin V, Adélaïde J, et al. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. Br J Haematol. 2009; 145:788–800. [PubMed: 19388938]

22. Shih AH, Abdel-Wahab O, Patel JP, Levine RL. The role of mutations in epigenetic regulators in myeloid malignancies. Nat Rev Cancer. 2012; 12:599–612. [PubMed: 22898539]

23. Chuman Y, Kurihashi W, Mizukami Y, Nashimoto T, Yagi H, Sakaguchi K. PPM1D430, a novel alternative splicing variant of the human PPM1D, can dephosphorylate p53 and exhibits specific tissue expression. J Biochem. 2009; 145:1–12. [PubMed: 18845566]

24. Ruark E, Snape K, Humburg P, et al. Mosaic PPM1D mutations are associated with predisposition to breast and ovarian cancer. Nature. 2013; 493:406–10. [PubMed: 23242139]

25. Kleiblova P, Shaltiel IA, Benada J, et al. Gain-of-function mutations of PPM1D/Wip1 impair the p53-dependent G1 checkpoint. J Cell Biol. 2013; 201:511–21. [PubMed: 23649806]

26. Akbari MR, Lepage P, Rosen B, et al. PPM1D mutations in circulating white blood cells and the risk for ovarian cancer. J Natl Cancer Inst. 2014; 106:djt323. [PubMed: 24262437]

27. Zhang L, Chen LH, Wan H, et al. Exome sequencing identifies somatic gain-of-function PPM1D mutations in brainstem gliomas. Nat Genet. 2014; 46:726–30. [PubMed: 24880341]

28. The, Cancer; Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013; 368:2059–74. Erratum, N Engl J Med 2013;369:98. [PubMed: 23634996]

29. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 2014; 505:495–501. [PubMed: 24390350]

30. Forbes SA, Bindal N, Bamford S, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011; 39:D945–D950. [PubMed: 20952405]

31. Kralovics R, Passamonti F, Buser AS, et al. A gain-of-function mutation of *JAK2* in myeloproliferative disorders. N Engl J Med. 2005; 352:1779–90. [PubMed: 15858187]

32. Russler-Germain DA, Spencer DH, Young MA, et al. The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. Cancer Cell. 2014; 25:442–54. [PubMed: 24656771]

33. Papaemmanuil E, Cazzola M, Boultwood J, et al. Somatic *SF3B1* mutation in myelodysplasia with ring sideroblasts. N Engl J Med. 2011; 365:1384–95. [PubMed: 21995386]

34. Campbell IM, Yuan B, Robberecht C, et al. Parental somatic mosaicism is underrecognized and influences recurrence risk of genomic disorders. Am J Hum Genet. 2014; 95:173–82. [PubMed: 25087610]

35. Chen T-C, Hou HA, Chou WC, et al. Dynamics of ASXL1 mutation and other associated genetic alterations during disease progression in patients with primary myelodysplastic syndrome. Blood Cancer J. 2014; 4:e177. [PubMed: 24442206]

36. Fasan A, Haferlach C, Alpermann T, et al. The role of different genetic subtypes of CEBPA mutated AML. Leukemia. 2014; 28:794–803. [PubMed: 24056881]

37. Martelli MP, Sportoletti P, Tiacci E, Martelli MF, Falini B. Mutational land-scape of AML with normal cytogenetics: biological and clinical implications. Blood Rev. 2013; 27:13–22. [PubMed: 23261068]

38. Kulasekararaj AG, Smith AE, Mian SA, et al. TP53 mutations in myelodysplastic syndrome are strongly correlated with aberrations of chromosome 5, and correlate with adverse prognosis. Br J Haematol. 2013; 160:660–72. [PubMed: 23297687]

39. Challen GA, Sun D, Jeong M, et al. Dnmt3a is essential for hematopoietic stem cell differentiation. Nat Genet. 2012; 44:23–31. [PubMed: 22138693]

40. Busque L, Patel JP, Figueroa ME, et al. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. Nat Genet. 2012; 44:1179–81. [PubMed: 23001125]
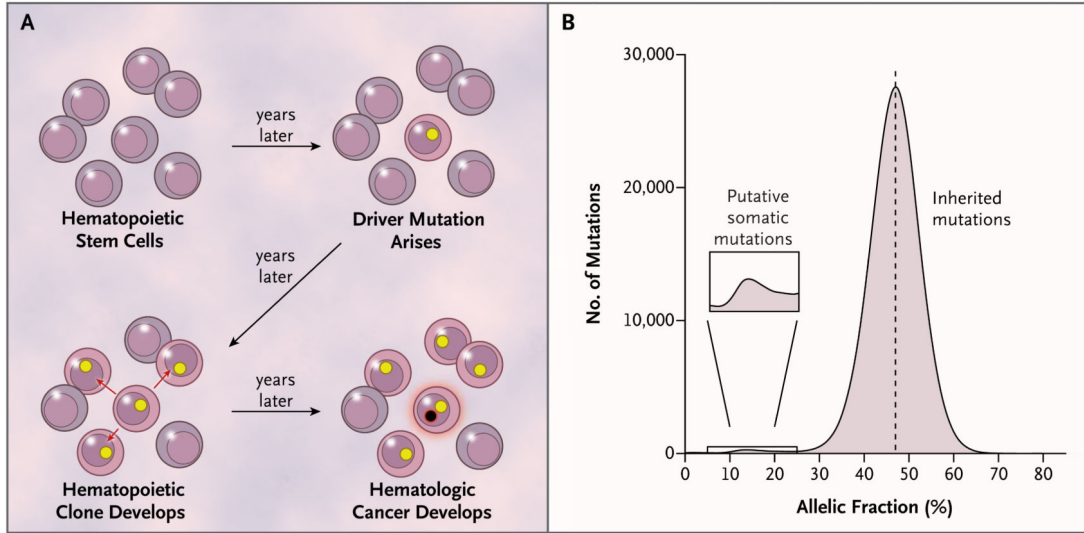
Figure 1. Clonal Expansion and Allelic Fractions

Panel A shows a model for the expansion of a single hematopoietic stem cell into a clonal population, under the influence of a somatic mutation (yellow circle), and the potential conversion of the clone into a hematologic cancer through subsequent mutation (black circle with red rim). Mutations present in the founder cell would be present at an appreciable allelic fraction (though <50%) in blood-derived genomic DNA. Panel B shows the distribution of allelic fractions observed in sequencing data for high-confidence, ultra-rare variants ascertained in the 12,380 study participants; the small bump at the left of this distribution represents putative somatic mutations.
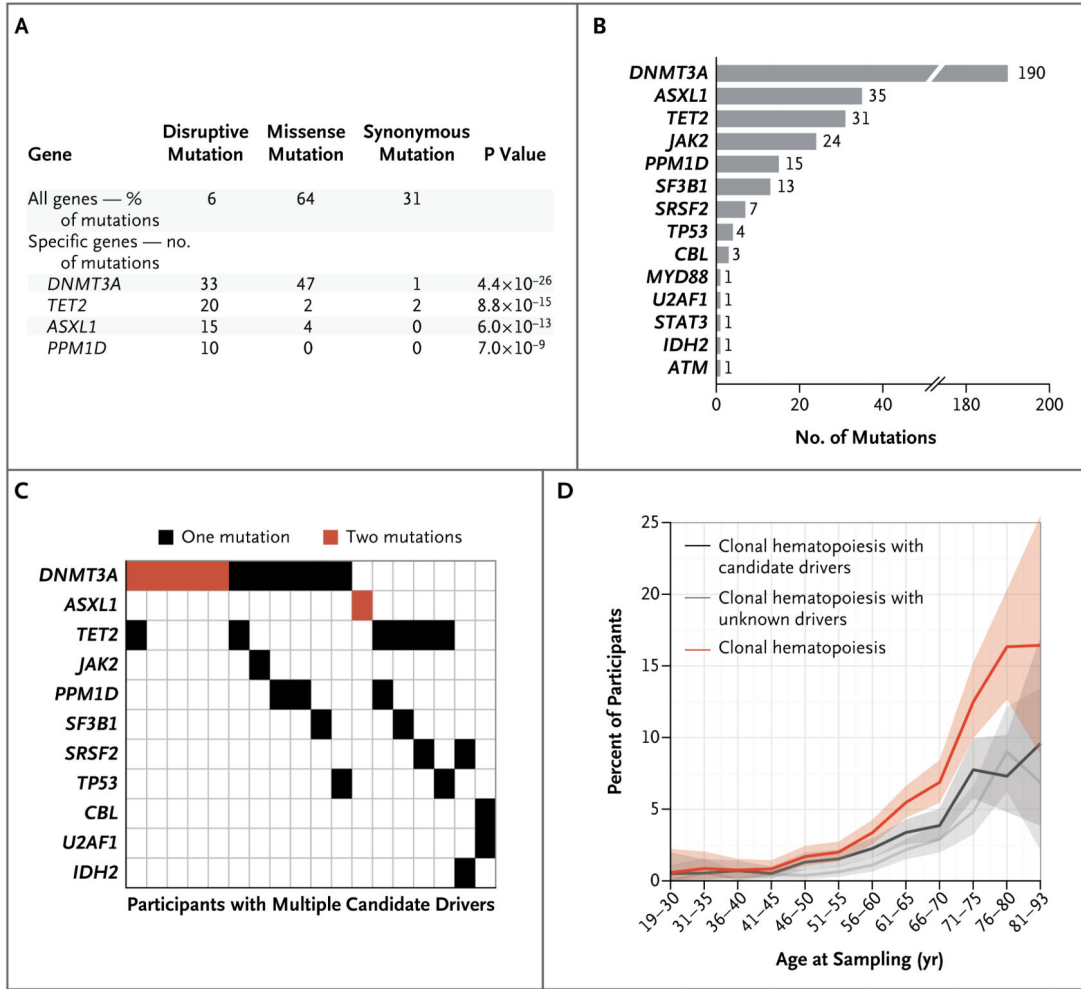
Figure 2. Candidate Driver Somatic Mutations

Panel A shows all genes identified as carrying a significant excess of disruptive (nonsense, frame-shift, and splice-site) somatic mutations among 11,845 participants with sequencing data of sufficient quality for the detection of somatic mutations. To control for potential systematic sequencing artifacts due to sequence context, somatic mutations observed in multiple participants were counted only once. Panel B shows the contribution of individual genes to the total number of candidate driver somatic mutations that were observed. Panel C shows a comutation plot for participants with multiple candidate driver somatic mutations. Panel D shows estimates for participants with clonal hematopoiesis with candidate drivers (carrying at least one candidate driver mutation), those with clonal hematopoiesis with unknown drivers (carrying three or more detectable somatic mutations and no candidate drivers), and those with clonal hematopoiesis and candidate or unknown drivers. The shaded bands represent 95% confidence intervals.
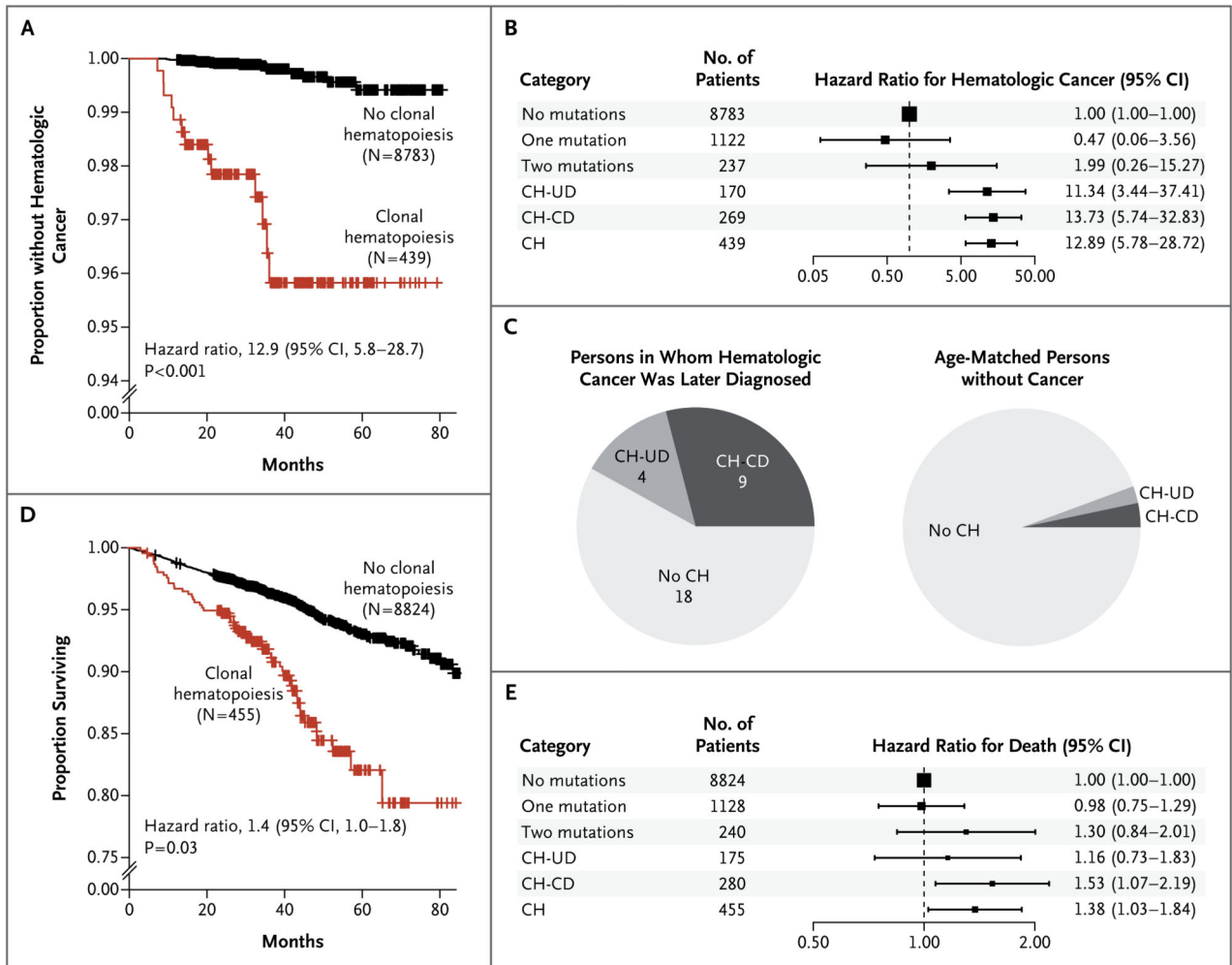
Figure 3. Risk of Hematologic Cancer for Participants with Clonal Hematopoiesis

Panels A and D show Kaplan–Meier plots of the proportions of participants who did not receive a diagnosis of hematologic cancer (Panel A) and for surviving participants (Panel D). Panels B and E show hazard ratios for hematologic cancer (Panel B) and death (Panel E) for participants with exactly one putative somatic mutation and no candidate drivers (one mutation), those with exactly two putative somatic mutations and no candidate drivers (two mutations), those with clonal hematopoiesis and unknown drivers (CH-UD), those with clonal hematopoiesis and candidate drivers (CH-CD), and those having clonal hematopoiesis with candidate or unknown drivers (CH), all compared with participants with no candidate drivers and no putative somatic mutations (no mutations). Panel C shows the proportions of participants who had clonal hematopoiesis with candidate or unknown drivers among 31 participants in whom hematologic cancers were diagnosed at least 6 months after DNA sampling, as compared with the proportions in a group of age-matched persons without hematologic cancer over a similar follow-up period.
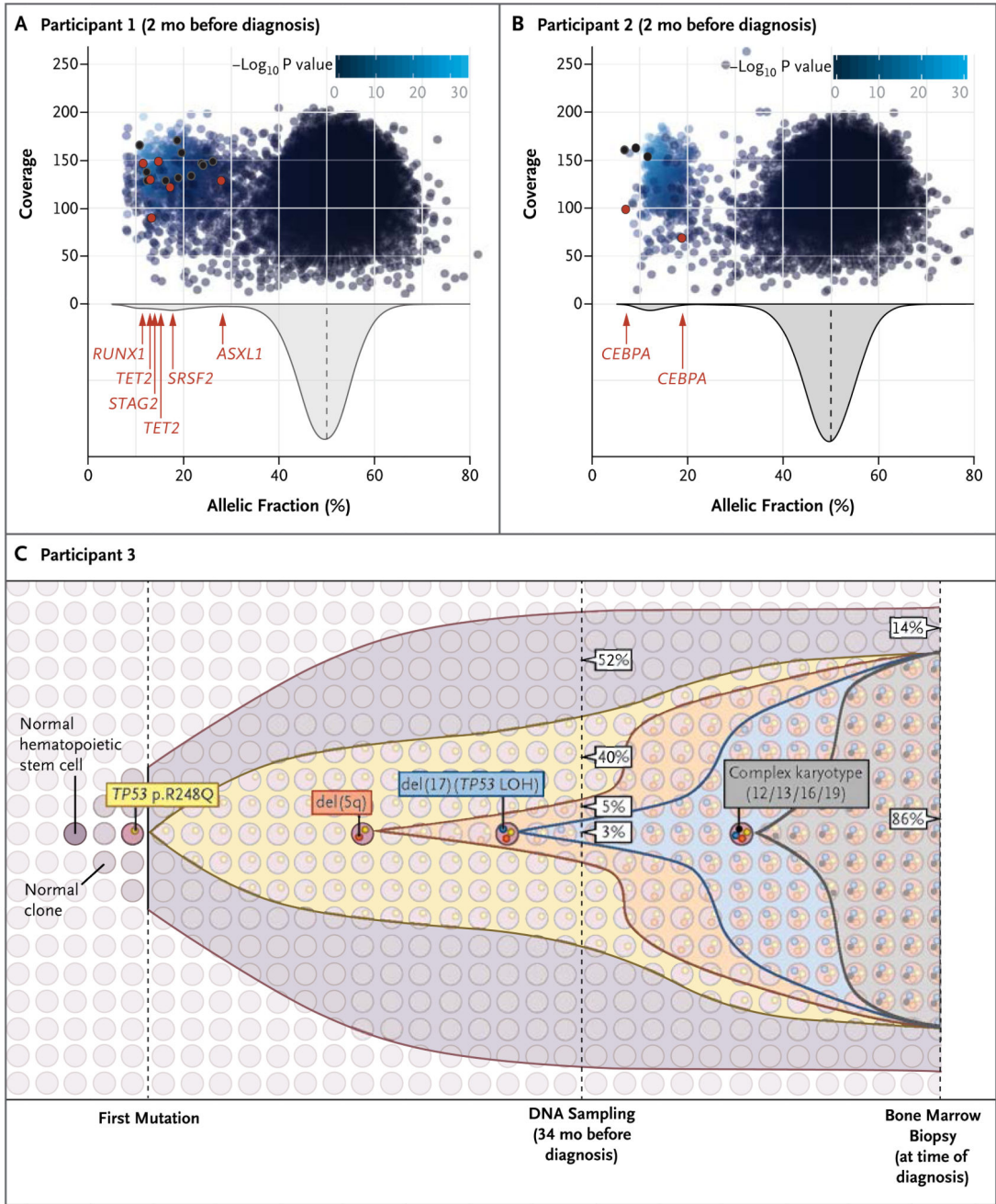
Figure 4. Hematopoietic Clones and Evolution in Three Patients Who Subsequently Received a Diagnosis of Myeloid Cancer

Panels A and B show the allelic fraction and coverage for rare heterozygous variants ascertained through whole-genome sequencing in Participant 1 and Participant 2, respectively, each of whom received a diagnosis of myeloid cancer 2 months after DNA sampling. Blue shading indicates the strength of evidence that a mutation was somatically acquired, with the negative $\log_{10}$ P value for the mutation being at an allelic fraction of less than 50% according to a binomial test. Mutations in black were initially ascertained through whole-exome sequencing. Mutations in red are candidate driver mutations. The histograms

show the overall distribution of allelic fractions, with the candidate driver mutations indicated in red. Panel C shows the progression from clonal hematopoiesis to myeloid cancer in Participant 3, in whom DNA was sampled 34 months before diagnosis and again at the time of diagnosis. Shading represents cell populations defined by specific combinations of mutations as shown; the percentages refer to the estimated representation of each cell population in the sample, at initial DNA sampling and then at the time of diagnosis (34 months later).