

SUPPLEMENTARY APPENDIX

Uncompromised ten-year survival of oldest old carrying somatic mutations in *DNMT3A* and *TET2*

TABLE OF CONTENTS

S1 LIST OF INVESTIGATORS	3
LEIDEN LONGEVITY STUDY	3
<i>Affiliations</i>	3
ROTTERDAM STUDY	3
<i>Affiliations</i>	3
GENOME OF THE NETHERLANDS CONSORTIUM	3
<i>Affiliations</i>	3
S2 STUDY POPULATION.....	5
ROTTERDAM STUDY	5
LEIDEN LONGEVITY STUDY	5
GENOME OF THE NETHERLANDS CONSORTIUM	5
S3 SEQUENCING AND VARIANT CALLING IN BLOOD-DERIVED DNA.....	6
ROTTERDAM STUDY	6
LEIDEN LONGEVITY STUDY & GENOME OF THE NETHERLANDS CONSORTIUM	6
S4 CALLING PUTATIVE SOMATIC MUTATIONS IN GENES LINKED TO CLONAL HEMATOPOIESIS	8
LIST OF 15 RECURRENTLY MUTATED GENES LINKED TO CLONAL OUTGROWTH	8
KNOWN HOTSPOT VARIANTS	8
RARE TRUNCATING VARIANTS	8
RARE MISSENSE VARIANTS IN <i>DNMT3A</i>	9
S5 SANGER SEQUENCING	10
S6 STATISTICAL ANALYSES.....	11
IBD COMPUTATIONS.....	11
PROSPECTIVE MORTALITY ANALYSIS.....	11
META-ANALYSIS	11
POWER CALCULATION	11
SUPPLEMENTARY TABLES.....	13
SUPPLEMENTARY TABLE S1: MUTATIONS IN THE RS ELDERLY SUBSAMPLE	13
SUPPLEMENTARY TABLE S2: MUTATIONS IN THE LLS ELDERLY SUBSAMPLE	15
SUPPLEMENTARY TABLE S3: BASELINE CHARACTERISTICS OF THE RS ELDERLY SUBSAMPLE	17
SUPPLEMENTARY TABLE S4: BASELINE CHARACTERISTICS OF THE LLS ELDERLY SUBSAMPLE	18
SUPPLEMENTARY TABLE S5: VARIANT DEFINITIONS	19
SUPPLEMENTARY TABLE S6: RARE <i>DNMT3A</i> MISSENSE MUTATIONS IN THE RS ELDERLY SUBSAMPLE	21
SUPPLEMENTARY TABLE S7: RARE <i>DNMT3A</i> MISSENSE MUTATIONS IN THE LLS ELDERLY SUBSAMPLE	22
SUPPLEMENTARY FIGURES	23
SUPPLEMENTARY FIGURE S1: OVERVIEW OF <i>TET2</i> MUTATIONS.....	23
SUPPLEMENTARY FIGURE S2-19: SANGER SEQUENCING RESULTS.....	24
REFERENCES.....	42

S1 LIST OF INVESTIGATORS

Leiden Longevity Study

Erik B. van den Akker^{1,2,*}, Steven J. Pitts³, Joris Deelen^{1,4}, Matthijs H. Moed¹, Shobha Potluri³, H. Eka D. Suchiman¹, Nico Lakenberg¹, Wesley J. de Dijcker¹, Anton J.M. de Craen⁵, Jeanine J. Houwing-Duistermaat⁶, David R. Cox^{3†}, Marian Beekman^{1,4}, Marcel J.T. Reinders², P. Eline Slagboom^{1,4}

Affiliations

- 1: Molecular Epidemiology, Leiden University Medical Center, Leiden, Netherlands
 - 2: The Delft Bioinformatics Lab, Delft University of Technology, Delft, Netherlands
 - 3: Rinat-Pfizer Inc, South San Francisco, United States of America
 - 4: Netherlands Consortium for Healthy Ageing, Leiden, Netherlands
 - 5: Gerontology and Geriatrics, Leiden University Medical Center, Leiden, Netherlands
 - 6: Medical Statistics, Leiden University Medical Center, Leiden, Netherlands
- *: Corresponding author: Erik Ben van den Akker; Leiden University Medical Center; Section of Molecular Epidemiology; Einthovenweg 20; 2300 RC, Leiden, The Netherlands; Building 2 [Research], Room T5-46; mail to: e.b.van_den_kker at lumc.nl
†: Deceased

Rotterdam Study

Jeroen van Rooij¹, André G. Uitterlinden^{1,2}, Robert Kraaij^{1,2}, Joyce B.J. van Meurs¹, Albert Hofman³

Affiliations

- 1) Internal Medicine, Erasmus Medical Center, Wytemaweg 80, 3015 CN, Rotterdam, The Netherlands.
- 2) Epidemiology, Erasmus Medical Center, Wytemaweg 80, 3015 CN, Rotterdam, The Netherlands.

Genome of the Netherlands Consortium

Analysis Group: Morris A. Swertz^{6,7} (Co-Chair), Laurent C. Francioli¹, Freerk van Dijk^{6,7}, Androniki Menelaou¹, Pieter B.T. Neerincx^{6,7}, Sara L. Pulit¹, Patrick Deelen^{6,7}, Clara C. Elbers¹, Pier Francesco Palamara², Itsik Pe'er^{2,8}, Abdel Abdellaoui⁹, Wigard P. Kloosterman¹, Mannis van Oven¹⁰, Martijn Vermaat¹¹, Mingkun Li¹², Jeroen F.J. Laros¹¹, Mark Stoneking¹², Peter de Knijff¹³, Manfred Kayser¹⁰, Jan H. Veldink¹⁴, Leonard H. van den Berg¹⁴, Heorhiy Byelas^{6,7}, Johan T. den Dunnen¹¹, Martijn Dijkstra^{6,7}, Najaf Amin¹⁵, K. Joeri van der Velde^{6,7}, Jouke Jan Hottenga⁹, Jessica van Setten¹, Elisabeth M. van Leeuwen¹⁵, Alexandros Kanterakis^{6,7}, Mathijs Kattenberg⁹, Lennart C. Karssen¹⁵, Barbera D.C. van Schaik¹⁶, Jan Bot¹⁷, Isaäc J. Nijman¹, David van Enckevort¹⁸, Hailiang Mei¹⁸, Vyacheslav Koval¹⁹, Kai Ye^{20,21}, Eric-Wubbo Lameijer²¹, Matthijs H. Moed²¹, Jayne Y. Hehir-Kwa²², Robert E. Handsaker^{5,23}, Shamil R. Sunyaev^{4,5}, Mashaal Sohail^{4,5}, Fereydoun Hormozdiari²⁴, Tobias Marschall²⁵, Alexander Schönhuth²⁵, Victor Guryev²⁶, Paul I.W. de Bakker^{1,3-5} (Co-Chair)

Cohort collection and sample management group: P. Eline Slagboom²¹, Marian Beekman²¹, Anton J.M. de Craen²¹, H. Eka D. Suchiman²¹, Albert Hofman¹⁵, Cornelia van Duijn¹⁵, Dorret I. Boomsma⁹, Gonneke Willemsen⁹, Bruce H. Wolffenbuttel²⁷, Mathieu Plattee¹⁶, Steven J. Pitts²⁸, Shobha Potluri²⁸, David R. Cox^{28,34}

Whole-genome sequencing: Qibin Li²⁹, Yingrui Li²⁹, Yuanping Du²⁹, Ruoyan Chen²⁹, Hongzhi Cao²⁹, Ning Li³⁰, Sujie Cao³⁰, Jun Wang^{29,31,32}; Ethical, Legal, and Social Issues: Jasper A. Bovenberg³³

Steering committee: Cisca Wijmenga^{6,7} (Principal Investigator), Morris A. Swertz^{6,7}, Cornelia M. van Duijn¹⁵, Dorret I. Boomsma⁹, P. Eline Slagboom²¹, Gertjan B. van Ommen¹¹, Paul I.W. de Bakker^{1,3-5}

Affiliations

- 1: Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands
- 2: Department of Computer Science, Columbia University, New York, NY, USA
- 3: Department of Epidemiology, University Medical Center Utrecht, Utrecht, The Netherlands
- 4: Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
- 5: Broad Institute of Harvard and MIT, Cambridge, MA, USA
- 6: Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
- 7: Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

- 8: Department of Systems Biology, Columbia University, New York, NY, USA
9: Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands
10: Department of Forensic Molecular Biology, Erasmus Medical Center, Rotterdam, The Netherlands
11: Leiden Genome Technology Center, Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
12: Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
13: Forensic Laboratory for DNA Research, Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
14: Department of Neurology, University Medical Center Utrecht, Utrecht, The Netherlands
15: Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands
16: Bioinformatics Laboratory, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam Medical Center, Amsterdam, The Netherlands
17: SURFsara, Science Park, Amsterdam, The Netherlands
18: Netherlands Bioinformatics Centre, Nijmegen, The Netherlands
19: Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands
20: The Genome Institute, Washington University, St. Louis, MO, USA
21: Section of Molecular Epidemiology, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands
22: Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
23: Department of Genetics, Harvard Medical School, Boston, MA, USA
24: Department of Genome Sciences, University of Washington, Seattle, WA, USA
25: Centrum voor Wiskunde en Informatica, Life Sciences Group, Amsterdam, The Netherlands
26: European Research Institute for the Biology of Ageing, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
27: Department of Endocrinology, University Medical Center Groningen, Groningen, The Netherlands
28: Rinat-Pfizer Inc, South San Francisco, CA, USA
29: BGI-Shenzhen, Shenzhen, China
30: BGI-Europe, Copenhagen, Denmark
31: Department of Biology, University of Copenhagen, Copenhagen, Denmark
32: The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark
33: Legal Pathways Institute for Health and Bio Law, Aerdenhout, The Netherlands
34: Deceased

S2 STUDY POPULATION

Rotterdam Study

The Rotterdam Study¹ is a population-based prospective cohort consisting of 14,926 participants aged over 45 years. All inhabitants of a suburb of the city of Rotterdam, Ommoord, were invited to participate. The objective of the Rotterdam Study is to investigate the determinants, incidence and progression of chronic disabling diseases in the elderly.

Exome sequencing data was created for a random subsample of 2,628 participants for whom height, weight and GWAS data was available. In total, 646 were aged over 80 years and were used for the current analyses (See Table S3 for baseline statistics). Selected participants were followed for mortality, with a median survival time of 8.7 years (range: 0.5-15.2 years) amongst the participants still alive. The Medical Ethical Committee of the Erasmus Medical Centre approved the study, and informed consent was obtained from all subjects.

Leiden Longevity Study

The Leiden Longevity Study² is a family-based study consisting of 421 Dutch Caucasian nonagenarian sibships (N=944) and for the current study the whole genome was sequenced for a subset of 218 unrelated nonagenarians (See Table S4 for baseline statistics). Families participating in the Leiden Longevity Study have at least two siblings meeting four inclusion criteria: (i) men are at least 89 years old and women are at least 91 years old, (ii) participants have at least one living brother or sister who fulfils the first criterion and is willing to participate, (iii) the nonagenarian sibship has an identical mother and father, and (iv) the parents of the nonagenarian sibship are Dutch and Caucasian. Participants were recruited between November 2006 and May 2008 and were followed for mortality thereafter, with a median survival time of 9.2 years amongst the participants still alive at the latest censoring in 2014 and a maximum of 11.0 years among the 218 sequenced nonagenarians. The Medical Ethical Committee of the Leiden University Medical Centre approved the study, and informed consent was obtained from all subjects.

Genome of the Netherlands Consortium

We employed sequencing data assayed on 98 individuals (Complete Genomics, >30x) of Dutch Caucasian origin aged below 65, collected by the Dutch Biobanking and Biomolecular Resources Research Infrastructure initiative (BBMRI-NL)³ for filtering variants observed in the elderly for either platform specific or population specific variant calls. Participants of BBMRI-NL are not selected for particular characteristics other than that they should reflect a random sample of the apparently healthy Dutch population.

S3 SEQUENCING AND VARIANT CALLING IN BLOOD-DERIVED DNA

Rotterdam Study

Exome sequencing data in the Rotterdam study was created, aligned and called in-house. In short, genomic DNA of Rotterdam Study participants were prepared from blood and fragmented into 200-400 bp fragments using Covaris Adaptive Focused Acoustics shearing according to the manufacturer's instructions (Covaris, Inc., Woburn, MA). Kapa Library preparation (Kapa Biosystems, Inc., Wilmington, MA) was performed on a Caliper Sciclone NGS workstation (Caliper Life Sciences, Hopkinton, MA), followed by exome capture of 6 samples per reaction using the Nimblegen SeqCap EZ V2 kit (Roche Nimblegen, Inc., Madison, WI). This capture targets 44Mb of exonic regions covering 30,246 coding genes, 329,028 exons and 710 miRNAs. Paired-end 2 x100 sequencing was performed on Illumina HiSeq2000 sequencer using Illumina TruSeq V3 chemistry (Illumina, Inc., San Diego, CA). Downstream analyses included demultiplexing (CASAVA software, Illumina), alignment to the hg19 (Genome Reference Consortium Human Reference 37) using the burrows-wheeler alignment⁴ (BWA) tool. Alignments were sorted by Picard (<http://broadinstitute.github.io/picard>) and subsequently processed by GATK⁵. Finally PCR duplicates were marked by Picard, Mean Depth of Coverage was determined using GATK, and Freemix values were estimated through verifyBAMid⁶. Samples that passed technical QC metrics were genotyped to gVCF level through GATKs Unified Genotyper. Samples were pooled by 300 per chromosome to generate a project-level vcf file using genotypeGVCFs. Indels and SNVs were filtered separately using GATKs Variant-Quality Score Recalibration. Following Genovese *et al.*⁷ and Jaiswal *et al.*⁸ we adjusted the standard GATK output to a lowered quality threshold (10 instead of 30), in order to be able to pick up low allele fraction events. Variants were filtered on a minimal read depth of 10 and a minimal alternative allele read depth of 3 for SNVs and 6 for indels, before being submitted to UCSC's Variant Annotation Integrator: <http://moma.ki.au.dk/genome-mirror/cgi-bin/hgVai>. Protein-altering variants annotated to reference transcripts of each gene (Supplementary Table S5) were maintained only.

Leiden Longevity Study & Genome of the Netherlands Consortium

Samples of the LLS were sent off for whole genome DNA sequencing and variant calling to Complete Genomics (>30x; Complete Genomics Inc., Mountain View California, version 1.3.0, <http://www.completemomics.com/customer-support/documentation>). Obtained data of the LLS was realigned and called to match the version of the processing pipeline employed for the 98 BBMRI samples (version 2.4.0.20) used as an ethnicity and platform matched control panel in this study. The thus obtained 218 mastervar files were merged and transformed to a VCF using the mkvcf function of cgatools (version 1.8.0.1) and filtered with in-house developed software employing the Rsamtools (version 1.20.5) and IRanges⁹ (version 2.2.9) packages for the statistical program R¹⁰ (version 3.2.1). Similar to GATK, the normal Complete Genomics preprocessing pipeline is tuned to deliver high quality genotyping calls of germ line variants, marked as 'VQHIGH' calls. In contrast, 'VQLOW' calls include variants called at lower genotyping quality threshold are now also included to increase the sensitivity for low allele fraction events. In short, using the merged VCF, 'AMBIGUOUS' calls were set to missing, whereas 'VQLOW' and 'VQHIGH' calls were maintained. Remaining multi-allelic loci were

removed, or in case when only a single variant allele was observed, were recoded to bi-allelic variants. Variants were filtered on a minimal read depth of 10 and a minimal alternative allele read depth of 3 for SNVs and 6 for indels, before being submitted to UCSC's Variant Annotation Integrator: <http://moma.ki.au.dk/genome-mirror/cgi-bin/hgVai>. Protein-altering variants annotated to reference transcripts of each gene (Supplementary Table S5) were maintained only.

S4 CALLING PUTATIVE SOMATIC MUTATIONS IN GENES LINKED TO CLONAL HEMATOPOIESIS

Somatic mutations are typically called by comparing the DNA sequence obtained from an affected (tumorous) tissue with an unaffected tissue. Variants that are present in the affected tissue, but are absent in the unaffected tissue, must have been acquired during life. In the absence of any control tissue, previous papers^{7,8} particularly stress the importance of a candidate approach when looking for putative somatic mutations linked to clonal expansion of hematopoietic stem cells. Where possible, we adopted their stringent inclusion criteria, to make our study maximally comparable to previously published studies on this topic.

List of 15 recurrently mutated genes linked to clonal outgrowth

We curated a list of genes previously reported to carry somatic mutations in at least 5 instances in at least one of three previous studies^{7,8,11} and includes the following 15 genes: *DNMT3A*, *TET2*, *ASXL1*, *TP53*, *JAK2*, *SF3B1*, *GNB1*, *CBL*, *SRSF2*, *GNAS*, *BRCC3*, *CREBBP*, *NRAS*, *RAD21*, *U2AF1* and *PPM1D*. These genes jointly harbour 98.1%, 98.8% and 88.2% of the somatic mutations identified by Xie *et al.*¹¹ (Figure 4a of their paper), Genovese *et al.*⁷ and Jaiswal *et al.*⁸ respectively.

Known Hotspot Variants

Adhering to the terminology of Xie *et al.*¹¹, we categorize putative somatic mutations in genes linked to clonal hematopoiesis as Known Hotspot Variants (KHKVs) or as Rare Truncating Variants (RTVs). For this we employ the gene-specific variant inclusion criteria set by Jaiswal *et al.*⁸, who curated a list of recurrently reported somatic mutations in hematological cancers in COSMIC¹². In the case of *PPM1D*, which was not amongst the initial 160 candidate genes of Jaiswal *et al.*⁸, but found to be significantly recurrently mutated by Genovese *et al.*⁷, we used the definitions of the latter. For the sake of completeness definitions for the 15 genes in our panel are repeated in Supplementary Table S5.

Rare Truncating Variants

In case a truncating variant was not listed amongst the gene specific variant definitions compiled by Jaiswal *et al.*⁸ (KHKVs), it could potentially match a broader term indicating a type of truncating event. In case the gene was annotated with one of such terms ('frameshift', 'nonsense' or 'splice-site') and the variant matched this definition, the following additional filters were applied:

1. **Filter for poor genomic regions:** Genovese *et al.*⁷ defined a set of filters aimed at false positive calls coming from genomic regions known to contain repetitive sequences or are otherwise associated with poor quality calls. This filter removed only one call from the LLS elderly subsample in RAD21_D123fs. For the sake of completeness we repeat the definitions of the filters defined by Genovese *et al.*⁷ here:
 - a. **Low complexity regions** in the 1000 Genomes Project phase 1¹³:
<https://github.com/lh3/varcmp/blob/master/scripts/LCR-hs37d5.bed.gz>
 - b. **Sites harboring markers failing Hardy Weinberg equilibrium** tests in the 1000 Genomes Project phase 1¹³:

- <https://github.com/lh3/varcmp/blob/master/scripts/1000g.hwebad.bed>
- c. **Sites with excess coverage** within the 1000 Genomes Project phase 1¹⁴.
 - d. **Segmental duplications** of the human genome^{15,16}:
<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/genomicSuperDups.txt.gz>
 - e. **Regions excluded from the strict mask** of the 1000 Genomes Project phase 1¹⁷ (inclusive filter! If overlapping *keep* the variant!):
ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/supporting/accessible_genome_masks/20120824_strict_mask.bed
2. **Max frequency:** A maximum of 1% within the cohort where observed in (RS: N=6, LLS: N=2), absence in the BBMRI control panel and a maximum of 0.1% in the Exome Variant Server¹⁸ (<http://evs.gs.washington.edu/EVS>) was required to pass the frequency filter.
3. **10%:** Following Jaiswal *et al.*⁸ we filtered for truncating variants occurring in the first or last 10% of the gene open reading frame, unless these have previously been reported to occur in hematological malignancies in these particular genes. None of the listed truncating variants met this criterion.

Rare Missense Variants in *DNMT3A*

Genovese *et al.*⁷ and Jaiswal *et al.*⁸ do not agree on the status of *DNMT3A* missense mutations as potential candidate driver mutations of clonal hematopoiesis. Whereas Genovese *et al.*⁷ include all missense mutations in *DNMT3A*, irrespective of frequency and VAF, Jaiswal *et al.*⁸ include only specific missense SNVs in *DNMT3A* in case they are recurrently reported in COSMIC¹². Using the additional filtering rules for RTVs, described above, we identified 10 and 11 carriers of *DNMT3A* missense mutations in respectively the RS and LLS elderly subsample (Supplementary Table S6 and S7). Since the VAFs of the *DNMT3A* missense mutations were also indicative for somatic events, we excluded these carriers from the control non-carrier group in the mortality analyses, unless these carriers also carried either a KHV or RTV, in which case they were considered a carrier. In addition, we repeated the mortality analyses including the *DNMT3A* missense mutations carriers as carriers of potential candidate driver mutations of clonal hematopoiesis.

S5 SANGER SEQUENCING

Bi-directional Sanger sequencing was performed to validate the presence of called mutations in our sequencing data. Only mutations observed in both forward and reverse reads were considered as being validated (Figure S2-S19).

S6 STATISTICAL ANALYSES

IBD computations

Genome-wide Identical By Descent (IBD) probabilities were inferred for siblings of mutant carriers using genotypes from GWAS chips, e.g. the Illumina 660Quad and Illumina OmniExpress, with MERLIN-0.10.2¹⁹. Data acquisition and pre-processing have been described previously²⁰. Siblings identical by descent for 2 alleles at the genetic location of the putative somatic mutations in *DNMT3A* and *TET2* were assumed to have inherited the same germ line DNA. Consequently, identified mutations observed in only one sibling in such regions are interpreted to be of somatic origin.

Prospective mortality analysis

Prospective mortality analyses of carriers versus non-carriers were performed using a left truncated Cox proportional hazards model adjusted for age at inclusion and sex using the Survival package²¹ of the statistical program R¹⁰:

$$\lambda(t) = \lambda_0(t) \times \exp(\beta_1 \times \text{age} + \beta_2 \times \text{sex} + \beta_3 \times \text{carrier} + \varepsilon) \quad \text{Equation 1}$$

Where the covariate *age* designates *age at inclusion*, and is provided in years, *sex* indicates the sex as either 1 (male) or 0 (female), and *carrier* whether the individual is carrying a mutation (1) or not (0). Kaplan Meier curves were plotted with R using follow-up time.

Meta-analysis

Fixed effect meta-analyses were performed using the meta package of the statistical program R¹⁰, using the Hazard Ratios and Standard Errors computed for the covariate indicating carriership versus non-carriership (β_3 Equation 1) in the RS and LLS elderly subsamples.

Power calculation

Power calculations were performed using the powerSurvEpi package of the statistical program R¹⁰ using the following information:

-THETA: (hypothesized effect size or Hazard Ratio): 1.454

Jaiswal *et al.* report a Hazard Ratio of 1.4 in 17,182 individuals; Genovese *et al.* report a Hazard Ratio of 1.53 (in those with a candidate driver (CH-CD)) in 12,300 individuals. A weighted mean yields theta = 1.454

-N: (number of individuals in our studies combined): 844

Remark 1: We now discover in both studies (RS: 646; LLS: 218) in the full candidate gene set.

Remark 2: Since Jaiswal *et al.* and Genovese *et al.* disagree on the status of *DNMT3A* missense mutation carriers, we exclude them (10 for both studies): 646-10+218-10=844

-alpha: (type I error rate): 0.05

-NCAR: (number of carriers): 79 (RS: 39; LLS: 40)

-NDEATH: (number of deceased individuals): 710

RS: 516; LLS: 194; excluding the 20 *DNMT3A* missense carriers

RESULTS (1): A power calculation adjusted for sex: 0.83, with a correlation between the outcome (mutation carriership) and covariate sex: $\rho^2 = 8.05 \times 10^{-6}$.

RESULTS (2): A power calculation adjusted for age: 0.81, with a correlation between the outcome (mutation carriership) and covariate age: $\rho^2 = 0.039$.

CONCLUSIONS: With a power of 0.81, we have sufficient power in our combined studies to detect an association between all-cause mortality and carriership of somatic mutations in case it is truly present.

SUPPLEMENTARY TABLES

SUPPLEMENTARY TABLE S1: MUTATIONS IN THE RS ELDERLY SUBSAMPLE

Gene	Location	Mutation	Age at inclusion [years]	Age at censoring [years]	Vital status (DEAD=1)	Read Depth (REF ALT)
<i>DNMT3A</i>	chr2:25471070	Q231*	90.1	94.3	1	32 12
	chr2:25471070	Q231*	88.8	92.9	0	28 07
	chr2:25464544	V657M	88.6	89.9	1	13 05
	chr2:25464544	V657M	81.5	87.6	1	34 05
	chr2:25463287	R736C	83.1	87.1	1	48 04
	chr2:25463184	S770W	87.2	94.7	1	20 06
	chr2:25463181	R771Q	83.5	93.2	1	08 07
	[A] chr2:25457242	R882H	89.0	99.3	1	08 04
	[B] chr2:25457242	R882H	81.7	96.2	1	19 14
	chr2:25457242	R882H	85.7	88.9	1	35 09
	chr2:25457242	R882H	80.4	89.7	0	09 04
	chr2:25457242	R882H	82.7	87.1	1	45 07
	chr2:25457243	R882C	88.5	93.7	1	33 06
	chr2:25457243	R882C	93.9	96.8	1	15 05
	chr2:25457243	R882C	82.4	88.2	1	38 03
<i>TET2</i>	[A] chr4:106155511	Q138*	89.0	99.3	1	26 11
	[C] chr4:106156417	Q440*	83.9	91.3	1	53 11
	chr4:106157329	Q744*	80.6	90.5	1	121 03
	chr4:106157384-					
	106157385	H762fs*5	89.2	89.8	1	18 16
	chr4:106164895-					
	106164896	Y1255*	82.9	93.9	1	30 06
	chr4:106180785	C1271*	81.4	92.3	0	84 03
	chr4:106180785	C1271*	86.0	93.5	1	35 11
	chr4:106193841	Q1435*	90.3	95.6	1	38 13
	chr4:106193849-					
	106193850	K1438fs*38	89.1	100.2	1	47 29
	[C] chr4:106196267	Q1534*	83.9	91.3	1	44 18
	chr4:106197458	E1931*	81.9	93.6	1	100 03
	chr4:106197458	E1931*	80.9	86.4	1	44 14
<i>ASXL1</i>	chr20:31021162	C387*	80.0	90.1	1	40 11
	chr20:31021550	Q517*	81.0	94.3	1	22 17
	chr20:31022264	W583*	89.1	100.1	1	22 05
<i>JAK2</i>	chr9:5073770	V617F	91.8	99.1	1	65 03
	chr9:5073770	V617F	89.6	91.8	1	31 13
	chr9:5073770	V617F	82.5	90.3	1	39 07
	chr9:5073770	V617F	89.1	92.1	1	32 04
<i>SF3B1</i>	[B] chr2:198266834	K700E	81.7	96.2	1	26 05
	chr2:198266834	K700E	92.0	96.6	1	21 04
	chr2:198266714	G740R	96.6	96.9	1	59 26

<i>BRCC3</i>	chrX:154305514	R89*	88.2	92.4	1	92 14
<i>RAD21</i>	chr8:117868887	S271*	82.1	91.9	0	29 05
<i>U2AF1</i>	chr21:44514777	Q157R	85.9	88.6	1	15 05
	chr21:44514777	Q157R	87.3	89.1	0	14 06
	chr21:44514777	Q157R	80.3	83.9	0	21 04

Table S1: Mutations identified in the RS elderly subsample (≥ 80 years) in 15 genes previously linked to clonal outgrowth of hematopoietic stem cells.

[A] These variants were observed in the same individual.

SUPPLEMENTARY TABLE S2: MUTATIONS IN THE LLS ELDERLY SUBSAMPLE

Gene	Location	Mutation	Age at inclusion [years]	Age at censoring [years]	Vital status (DEAD=1)	Read Depth (REF ALT)
DNMT3A	chr2:25470498	R326C	99.2	99.6	1	29 03
	chr2:25469991	C351fs*54	93.7	94.9	1	#¶34 21
	chr2:25469922	Q374*	92.5	97.5	1	§18 04
	[E] chr2:25468922-					
		V480fs*9	97.5	103.6	1	#¶30 08
	[F] chr2:25468154					
		L508fs*141	92.2	100.3	1	#¶34 10
	[F] chr2:25467432					
		M548I	94.3	99.1	1	12 12
	[G] chr2:25467083					
		R598*	91.4	100.0	0	08 03
TET2	chr2:25463578	D702Y	93.0	101.0	1	33 10
	[G] chr2:25463296-					
		E733*	96.5	107.0	0	¶47 17
	[H] chr2:25463248					
		R749C	97.0	99.3	1	54 23
	[H] chr2:25463182					
		R771*	95.6	99.9	1	¶46 09
	[H] chr2:25463182					
		R771*	94.2	99.1	1	¶57 08
	[I] chr2:25457243					
		R882C	97.2	104.4	1	09 03
	[I] chr2:25457243					
		R882C	101.1	101.4	1	¶08 04
ASXL1	chr4:106155737	V213fs*35	92.1	97.8	1	#¶30 11
	chr4:106155766	G223fs*25	92.8	93.5	1	¶60 09
	[G] chr4:106156132					
		K345*	102.2	102.4	1	04 32
	[G] chr4:106156686					
		C529*	93.9	95.3	1	¶23 10
	[H] chr4:106156758-					
		P554fs*11	94.1	100.1	1	¶33 10
	[H] chr4:106157246-					
		S716fs*5	93.7	99.8	1	¶24 09
	[H] chr4:106157782					
		V895fs*24	94.1	100.1	1	¶36 13
TP53	chr4:106157914	Q939*	91.0	92.7	1	#¶26 05
	chr4:106158108	W1003*	97.5	104.2	1	¶43 06
	chr4:106158256	Q1053*	90.0	99.9	0	44 03
	[I] chr4:106180916					
		D1315fs*46	96.0	97.8	1	57 06
	[I] chr4:106196213					
		R1516*	91.05	95.7	1	¶42 08
	[I] chr4:106196222					
		G1519*	94.3	98.3	1	#¶44 07
	chr4:106197353	R1896fs*10	88.9	93.5	1	¶47 22
SF3B1	[I] chr20:31021250					
		R417*	96.0	97.8	1	14 03
	[G] chr20:31021586					
		Q529*	93.7	96.3	1	24 22
	chr20:31022288	Y591*	91.9	92.1	1	17 08
[F] chr2:198267360	chr20:31022716	R734fs*8	92.9	98.8	1	23 22
	[G] chr20:31023418-					
		S968fs*7	102.2	102.4	1	12 11
	[F] chr2:198266834					
		L35fs*6	93.4	99.8	1	33 07
	chr17:7577042	L299fs*44	97.1	102.2	1	34 06
	chr2:198266834	K666R	97.3	101.4	1	41 13
	chr2:198266834	K700E	97.3	101.4	1	10 03
	chr2:198266834	K700E	97.0	99.3	1	09 05

<i>GNB1</i>	chr1:1747229	K57E	91.8	97.8	1	43 13
	chr1:1747229	K57E	96.0	96.6	1	42 10
	chr1:1747229	K57E	100.5	102.2	1	26 07
<i>U2AF1</i>	chr21:44524456	S34F	94.5	96.8	1	28 13
<i>PPM1D</i> ^[E]	chr17:58740438-58740439	N448fs*6	97.5	103.6	1	42 26
	chr17:58740526	N477fs*4	92.6	95.5	1	39 08
	chr17:58740529	C478*	95.1	100.4	1	54 04
	chr17:58740665	Q524*	91.7	94.9	1	41 19

Table S2: Mutations identified in an elderly subsample of the Leiden Longevity Study (≥ 89 years) in 15 genes previously linked to clonal outgrowth of hematopoietic stem cells.

[E] These variants were observed in the same individual.

¶ Variants in *TET2* and *DNMT3A* validated by Sanger sequencing.

§ Variant in *DNMT3A* that could not be confirmed by Sanger sequencing.

Variants in *TET2* and *DNMT3A* confirmed to be somatic. Sanger sequencing in siblings of mutation carriers who were Identical By Descent (IBD2) for the regions of interest confirmed the somatic origin.

SUPPLEMENTARY TABLE S3: BASELINE CHARACTERISTICS OF THE RS ELDERLY SUBSAMPLE

Total # individuals >=80 years	646 (33.9% male)
Age at inclusion [years]	83.5 (80.0-105.8)##
Follow-up time [years]	8.7 (0.5-15.2)##
Number of deaths [N, %]	525 (81.3%)
Age at censoring [years]	90.0 (81.2-108.2)##

Table S3: Baseline characteristics of the RS elderly subsample, consisting of 646 unrelated elderly participants of age ≥80 years of the Rotterdam Study, for whom exome sequencing data was available.

Listed are median values and the range.

SUPPLEMENTARY TABLE S4: BASELINE CHARACTERISTICS OF THE LLS ELDERLY SUBSAMPLE

Total # individuals >=89 years	218 (37.6% male)
Age at inclusion [years]	93.7 (88.9-103.4)##
Follow-up time [years]	9.2 (8.5-11.0)##
Number of deaths [N, %]	204 (93.6%)
Age at censoring [years]	97.8 (90.0-107.0)##

Table S4: Baseline characteristics of the LLS elderly subsample, consisting of 218 unrelated elderly participants of age ≥89 years of the Leiden Longevity Study, for whom whole genome sequencing data was available.

Listed are median values and the range.

SUPPLEMENTARY TABLE S5: VARIANT DEFINITIONS

GENE [TRANSCRIPT]	INCLUSION CRITERIA
<i>DNMT3A</i> [NM_022552]	RTV: frameshift nonsense splice-site KHV: P307S P307R R326H R326L R326C R326S R366P R366H R366G A368T F414L F414S F414C C497Y Q527H Q527P Y533C G543A G543S G543C L547H L547P L547F M548I M548K G550R W581R W581G W581C G646V G646E L653W L653F V657A V657M R659H Y660C R676W R676Q G685R G685E G685A D686Y D686G G699R G699S G699D P700S P700R P700Q D702N D702Y V704M V704G I705F I705T I705S C710S S714C N717S N717I P718L R720H R720G Y724C R729Q R729W R729G F731L F732del F732S F732L F734L F734C Y735C Y735N Y735S R736H R736C R736P L737H L737V L737F L737R A741V R749C R749L F751L F752del F752C F752L F752I F752V L754R L754H F755S F755I F755L M761I M761V G762C S770W S770P R771Q F772I F772V L773R E774K E774D D781G R792H G796D G796V N797Y N797H P799R P799H R803S P804S P804L S828N K829R Q842E P849L D857N W860R F868S G869S G869V M880V S881R S881I R882H R882P R882C R882G Q886R G890D L901R L901H P904L F909C A910P
<i>TET2</i> [NM_001127208]	RTV: frameshift nonsense splice-site KHV: S282F N312S L346P S460F D666G P941S C1135Y
<i>ASXL1</i> [NM_015338]	RTV: frameshift nonsense (only in exon 11-12) KHV: none
<i>TP53</i> [NM_001126112]	RTV: frameshift nonsense splice-site KHV: S46F G105C G105R G105D G108S G108C R110L R110C T118A T118R T118I S127F S127Y L130V L130F K132Q K132E K132W K132R K132M K132N C135W C135S C135F C135G C135Y Q136K Q136E Q136P Q136R Q136L Q136H A138P A138V A138A A138T T140I C141R C141G C141A C141Y C141S C141F C141W V143M V143A V143E L145Q W146C W146L L145R V147G P151T P151A P151S P151H P151R P152S P152R P152L T155P V157F R158H R158L A159V A159P A159S A159D A161T A161D Y163N Y163H Y163D Y163S Y163C K164E K164M K164N K164P H168Y H168P H168R H168L H168Q M169I M169T M169V E171K E171Q E171G E171A E171V E171D V172D V173M V173L V173G R174W R175G R175C R175H C176R C176G C176Y C176F C176S P177R P177T P177L H178P H178Q H179Y H179R H179Q R181C R181Y D186G G187S P190L P190T H193N H193P H193L H193R L194F L194R I195F I195N I195T V197L G199V Y205N Y205C Y205H D208V R213Q R213P R213L R213Q H214D H214R S215G S215I S215R V216M V217G Y220N Y220H Y220S Y220C E224D I232F I232N I232T I232S Y234N Y234H Y234S Y234C Y236N Y236H Y236C M237V M237K M237I C238R C238G C238Y C238W N239T N239S S241Y S241C S241F C242G C242Y C242S C242F G244S G244C G244D G245S G245R G245C G245D G245A G245V G245S M246V M246K M246R M246I N247I R248W R248G R248Q R249G R249W R249T R249M P250L I251N L252P I254S I255F I255N I255S L257Q L257P E258K E258Q D259Y S261T G262D G262V L265P G266R G266E G266V R267W R267Q R267P E271K V272M V272L R273S R273G R273C R273H R273P R273L V274F V274D V274A V274G V274L C275Y C275S C275F A276P C277F P278T P278A P278S P278H P278R P278L G279E R280G R280K R280T R280I R280S D281N D281H D281Y D281G D281E R282G R282W R282Q R282P E285K E285V E286G E286V E286K K320N L330R G334V R337C R337L A347T L348F T377P
<i>JAK2</i> [NM_004972]	RTV: none KHV: N533D N533Y N533S H538R K539E K539L I540T I540V V617F R683S R683G deI/ins537_539L del/ins538_539L del/ins540_543MK del/ins540_544MK del/ins541_543K del542_543 del543_544 ins11546_547
<i>SF3B1</i>	RTV: none

[NM_012433]	KHV: G347V R387W R387Q E592K E622D Y623C R525L R625C H662Q H662D K666N K666T K666E K666R K700E V701F A708T G740R G740E A744P D781G E783K
<i>GNB1</i> [NM_002074]	RTV: none KHV: K57N K57M K57E K57T I80T I80N
<i>CBL</i> [NM_005188]	RTV: none KHV: RING_finger_missense_p.381_421
<i>SRSF2</i> [NM_003016]	RTV: none KHV: Y44H P95H P95L P95T P95R P95A P107H P95fs
<i>GNAS</i> [NM_016592]	RTV: none KHV: R201(844)S R201(844)C R201(844)H R201(844)L Q227(870)K Q227(870)R Q227(870)L Q227(870)H R374(1017)C
<i>BRCC3</i> [NM_024332]	RTV: frameshift nonsense splice-site KHV: none
<i>CREBBP</i> [NM_004380]	RTV: frameshift nonsense splice-site KHV: D1435E R1446L R1446H R1446C Y1450C P1476R Y1482H H1487Y W1502C Y1503D Y1503H Y1503F S1680del
<i>NRAS</i> [NM_002524]	RTV: none KHV: G12S G12R G12C G12N G12P G12Y G12D G12A G12V G12E G13S G13R G13C G13N G13P G13Y G13D G13A G13V G13E G60E G60R Q61R Q61L Q61K Q61P Q61H Q61Q
<i>RAD21</i> [NM_006265]	RTV: frameshift nonsense splice-site KHV: R65Q H208R Q474R
<i>U2AF1</i> [NM_006758]	RTV: none KHV: D14G S34F S34Y R35L R156H R156Q Q157R Q157P
<i>PPM1D*</i> [NM_003620]	RTV: frameshift nonsense splice-site KHV: none

Table S5: Gene-specific variant inclusion criteria set by Jaiswal *et al.* for the 15 genes in our panel.

RTV: Rare Truncating Variant: a non-specific annotation indicating a type of truncating event, i.e.: ‘frameshift’, ‘nonsense’, ‘splice-site’. Recurrent truncating events in hematological malignancies are annotated as KHV.

KHV: Known Hotspot Variant: Variants recurrently reported in COSMIC¹² to be mutated in hematological malignancies, as curated by Jaiswal *et al.*⁸

*: No definition provided by Jaiswal *et al.*⁸, instead a definition set by Genovese *et al.*⁷ was employed.

SUPPLEMENTARY TABLE S6: RARE *DNMT3A* MISSENSE MUTATIONS IN THE RS ELDERLY SUBSAMPLE

Gene	Location	Mutation	Age at inclusion [years]	Age at censoring [years]	Vital status (DEAD=1)	Read Depth (REF ALT)
<i>DNMT3A</i>	chr2:25471061	G234R	85.2	88.4	1	17 13
	chr2:25469073	A462V	81.4	97.6	1	29 23
	chr2:25469037	R474H	83.3	86.7	1	42 35
	chr2:25466797	V636M	80.4	86.8	1	29 13
	chr2:25466790	S638C	80.7	84.8	0	25 60
	chr2:25463248	R749S	81.5	83.3	1	29 30
	chr2:25463248	R749S	87.5	89.5	1	20 13
	chr2:25459805	K826N	81.0	88.0	1	15 40
	chr2:25458661	N838D	81.5	86.4	1	193 30
	chr2:25458661	N838D	84.5	89.3	1	46 17

Table S6: Rare *DNMT3A* missense mutations identified in an elderly subsample of the Rotterdam Study¹ (≥ 80 years)

SUPPLEMENTARY TABLE S7: RARE DNMT3A MISSENSE MUTATIONS IN THE LLS ELDERLY SUBSAMPLE

Gene	Location	Mutation	Age at inclusion [years]	Age at censoring [years]	Vital status (DEAD=1)	Read Depth (REF ALT)
<i>DNMT3A</i>	chr2:25471010	T251P	92.7	94.8	1	12 03
*	chr2:25470605	F290S	88.9	93.5	1	45 17
	chr2:25470554	P307L	95.8	99.7	1	41 06
	chr2:25470465	S337P	96.9	98.0	1	42 14
	chr2:25470465	S337P	91.6	98.7	1	38 06
	chr2:25469119	A447T	89.8	94.4	1	19 05
	chr2:25466805	P633L	94.7	97.8	1	20 05
	chr2:25464535	Y660N	95.3	95.6	1	34 05
	chr2:25462016	N797K	93.9	97.7	1	25 04
	chr2:25458578	E865D	92.0	96.0	1	36 07
	chr2:25457183	F902V	95.5	97.8	1	30 23

Table S7: Rare *DNMT3A* missense mutations identified in an elderly subsample of the Leiden Longevity Study² (≥ 89 years).

* The carrier of this variant is also carrying *TET2*_R1896fs*10 (Supplementary Table S2).

SUPPLEMENTARY FIGURES

SUPPLEMENTARY FIGURE S1: OVERVIEW OF TET2 MUTATIONS

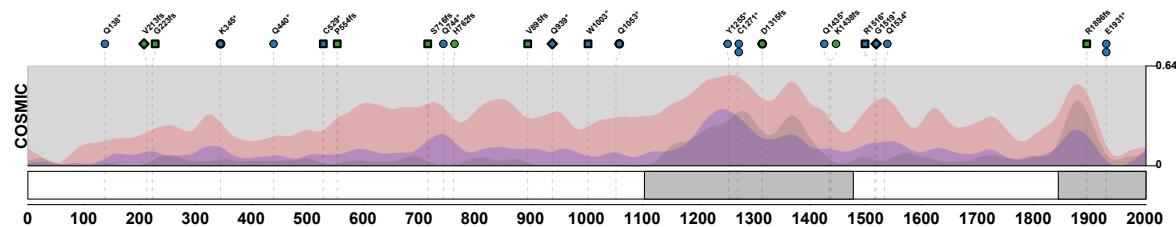


Figure S1: Overview of mutations in *TET2* identified in the RS and LLS elderly

subsample: Variants are annotated at the top with a color-coding to indicate the impact and a shape to indicate the types of follow-up experiments. Circles indicate mutations detected in our sequencing data; squares indicate mutations also validated by Sanger sequencing; diamonds indicate mutations also validated by Sanger sequencing and absent in an IBD2 matched sib, i.e. confirming somatic variations. Mutations identified in multiple carriers are indicated with stacked annotations and those having bold borders were identified in the LLS. Missense variants are only included whenever they are present on a curated list of recurrently reported variants in Catalogue Of Somatic Mutations In Cancer (COSMIC)¹² assembled by Jaiswal *et al.*⁸. Domains: The conserved domains (grey) in the protein are taken from Langemeijer *et al.*²². COSMIC: Densities of somatic variants identified in hematopoietic or lymphoid tissue collected by the Catalogue Of Somatic Mutations In Cancer (COSMIC)¹² database: all small variants (red), missense SNVs (grey), small variants confirmed to be of somatic origin (blue).

SUPPLEMENTARY FIGURE S2-19: SANGER SEQUENCING RESULTS

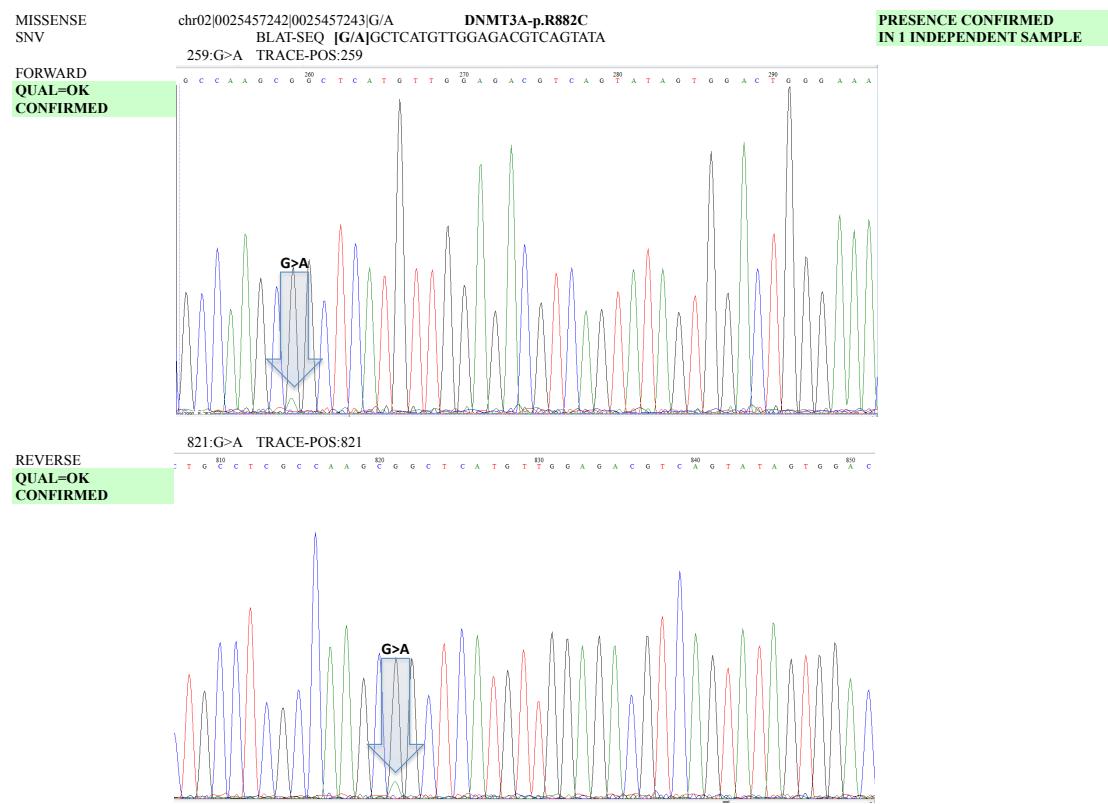


Figure S2: Sanger sequencing results of *DNMT3A* R882C

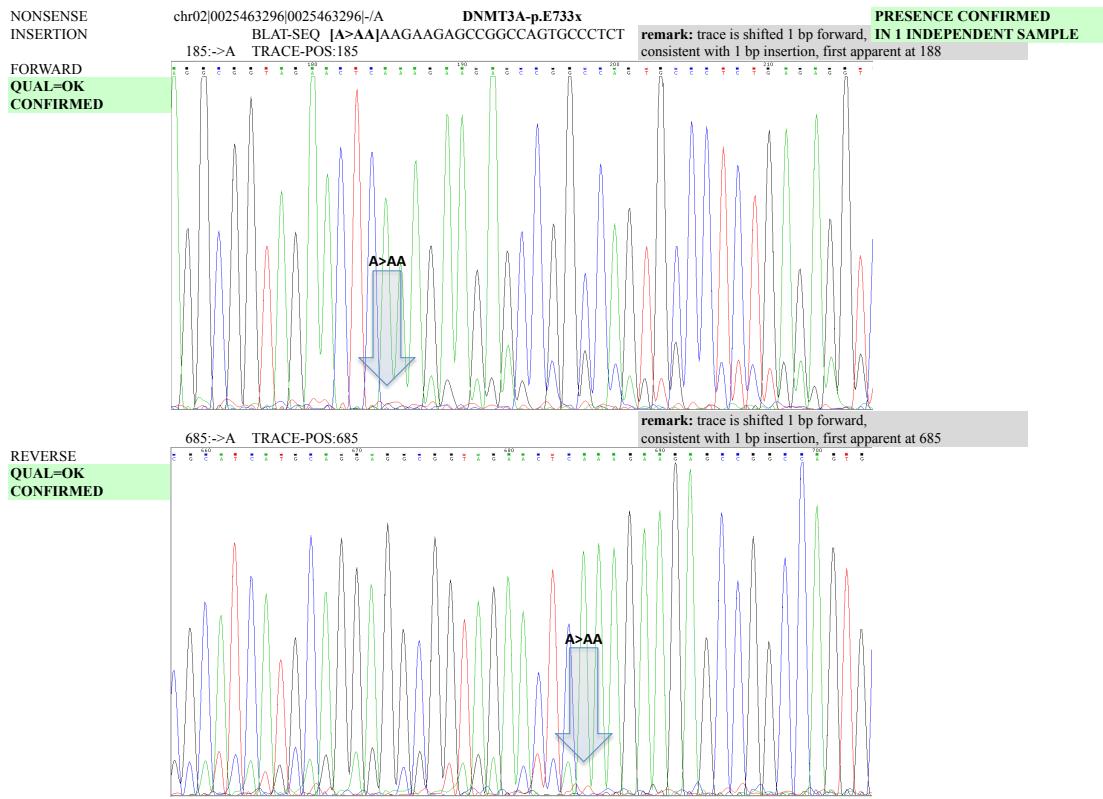


Figure S3: Sanger sequencing results of *DNMT3A* E734*

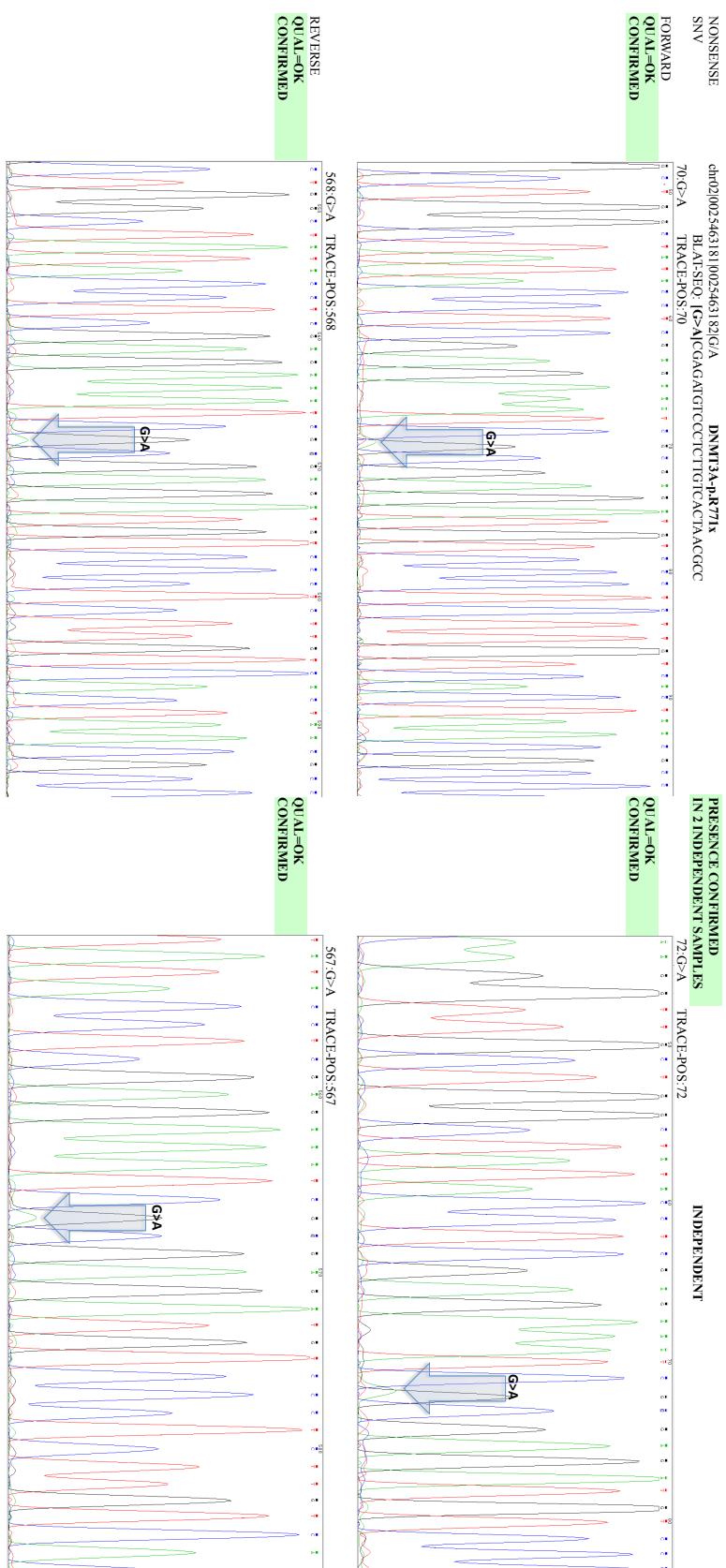


Figure S4: Sanger sequencing results of *DNMT3A* R771*

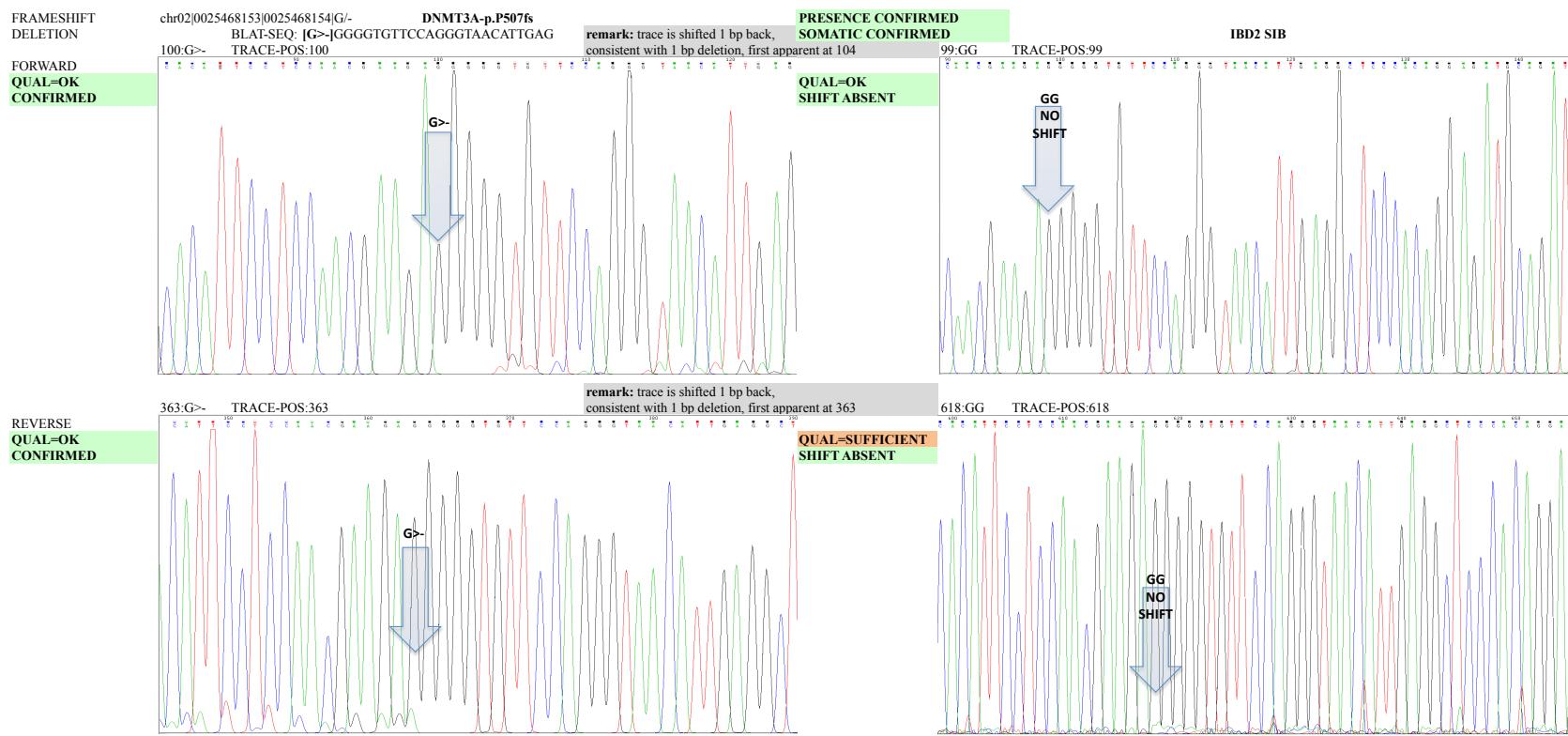


Figure S5: Sanger sequencing results of *DNMT3A* P508fs

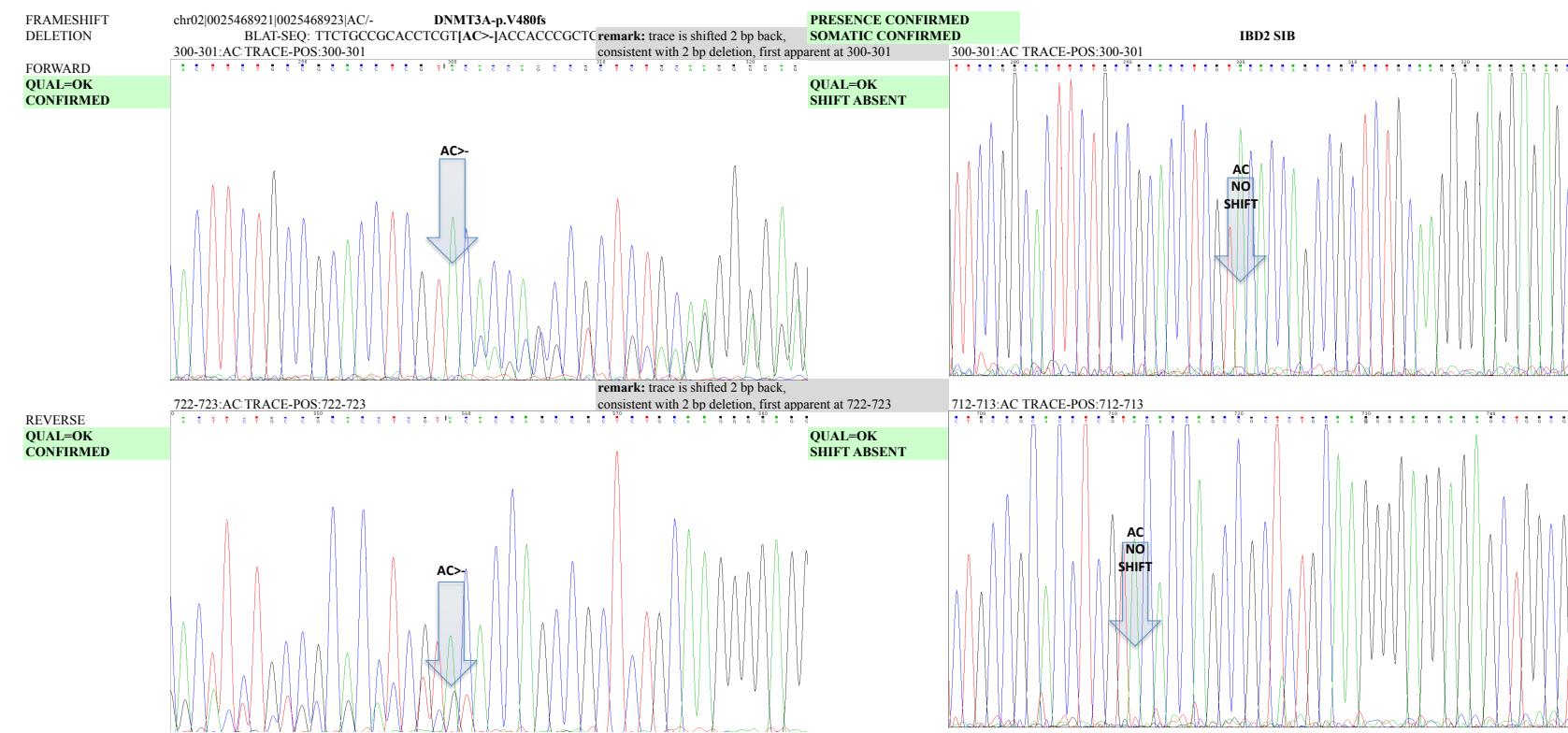


Figure S6: Sanger sequencing results of *DNMT3A* V480fs

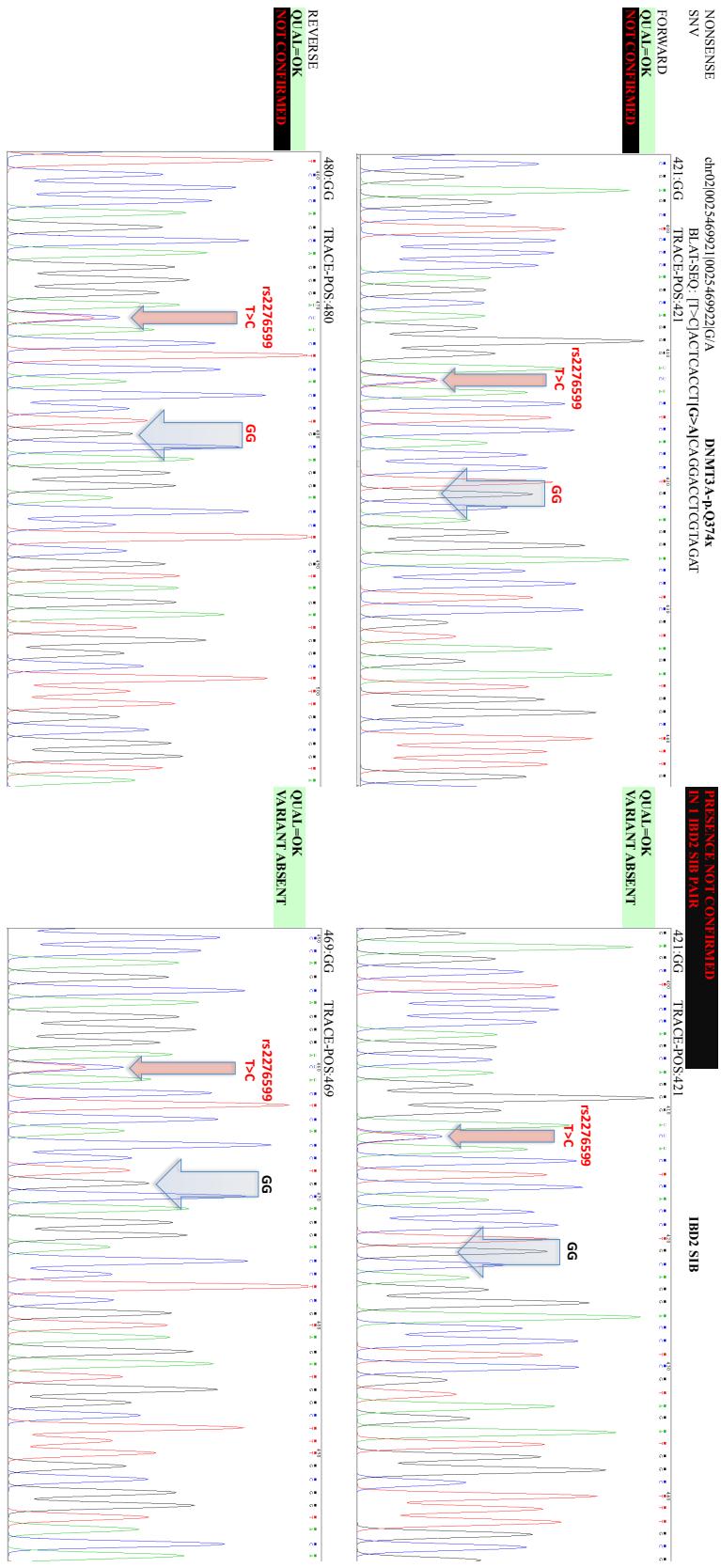
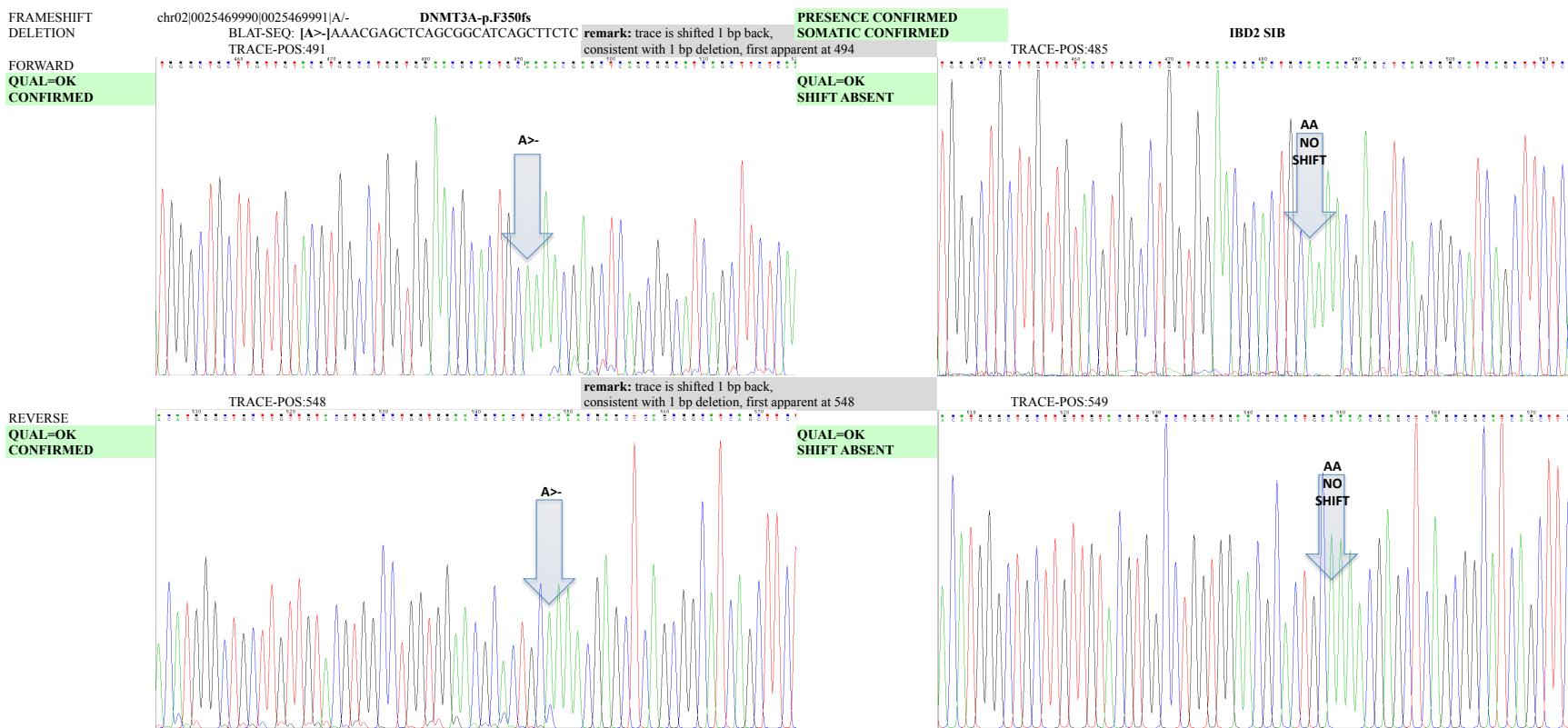


Figure S7: Sanger sequencing results of *DNMT3A* p.Q374*

Figure S8: Sanger sequencing results of DNMT3A F351fs



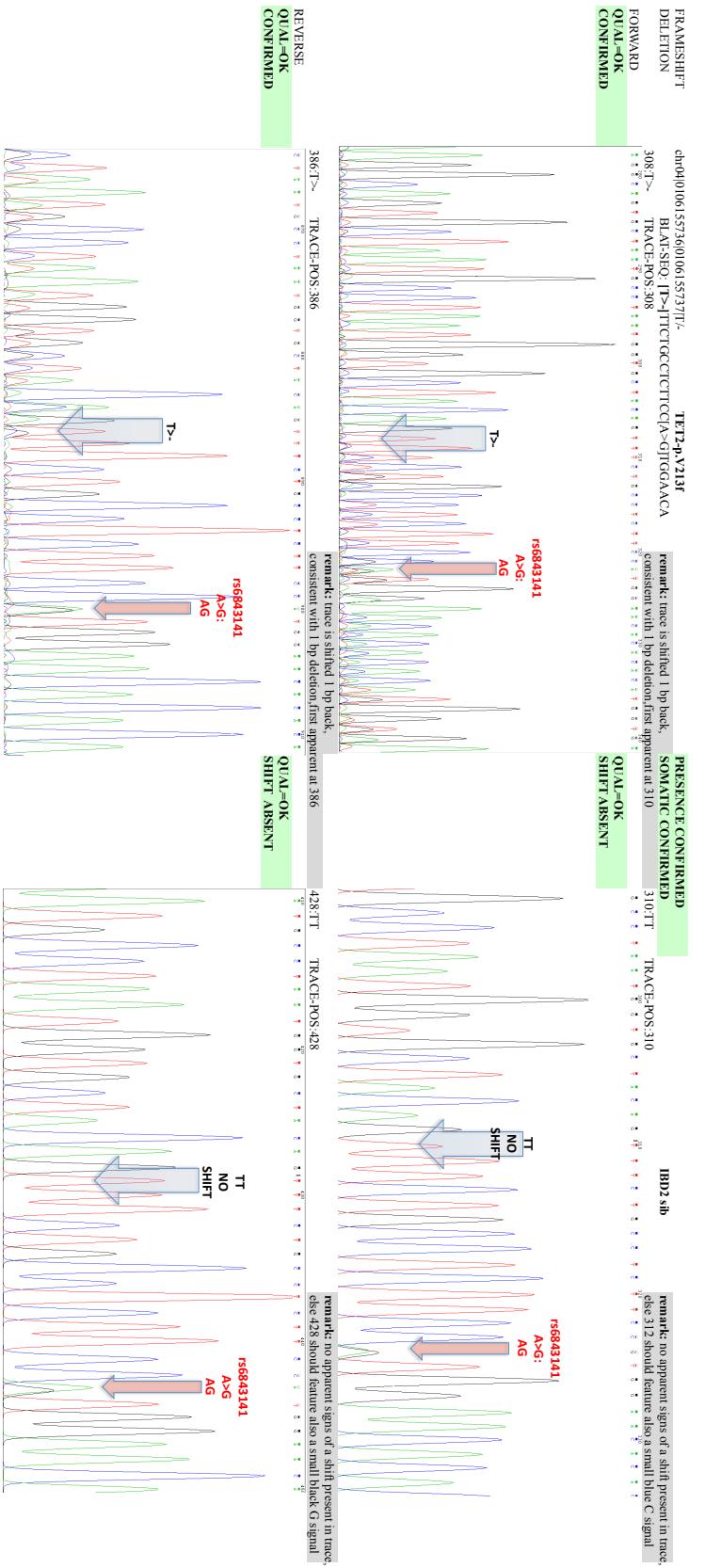


Figure S9: Sanger sequencing results of *TET2* V213fs

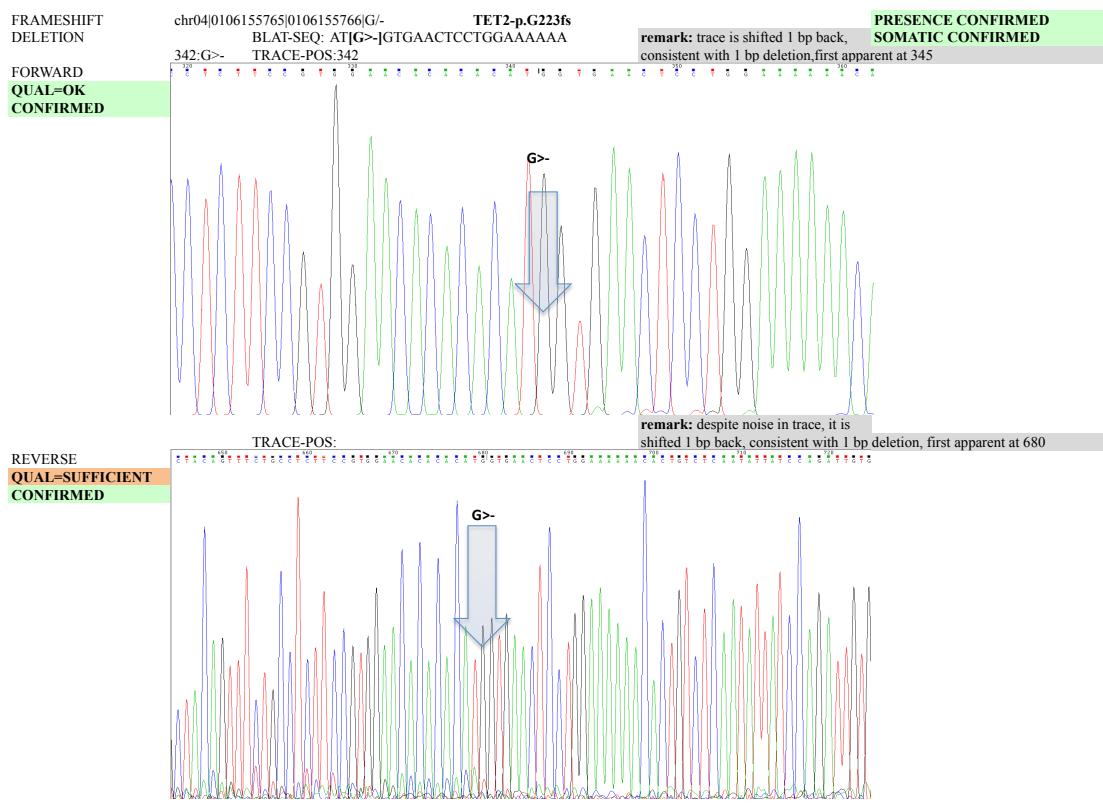


Figure S10: Sanger sequencing results of *TET2* G223fs

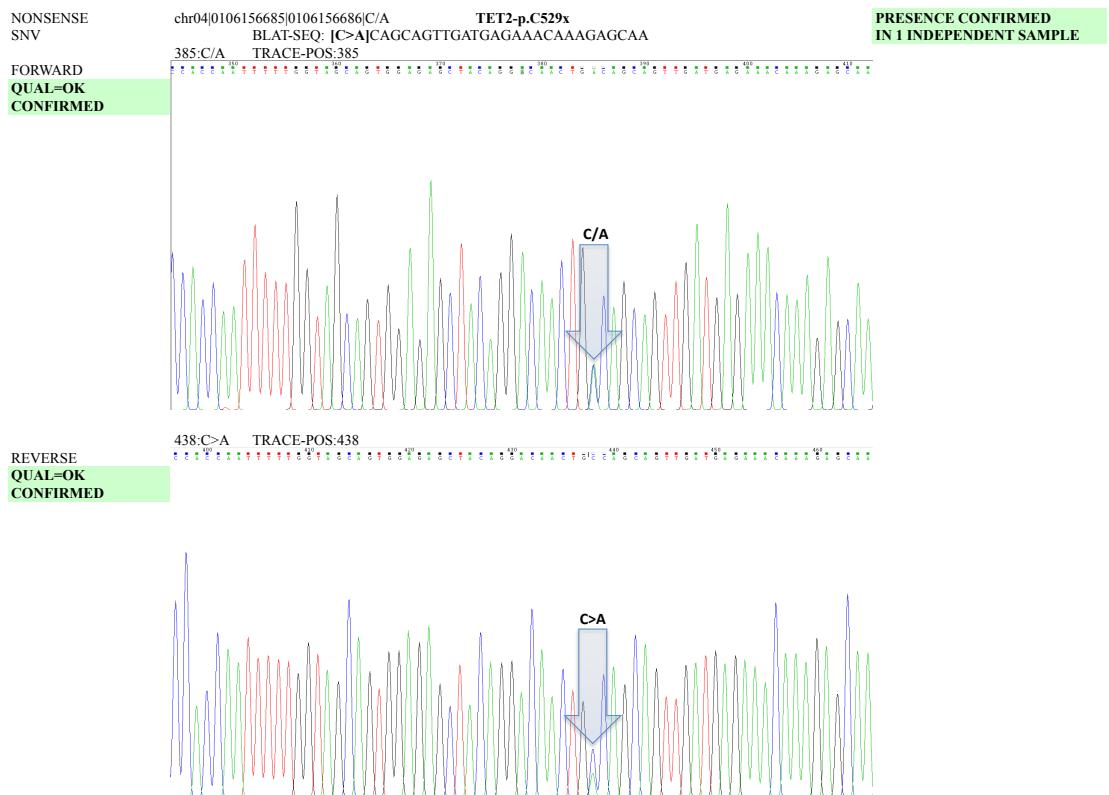


Figure S11: Sanger sequencing results of *TET2* C529*

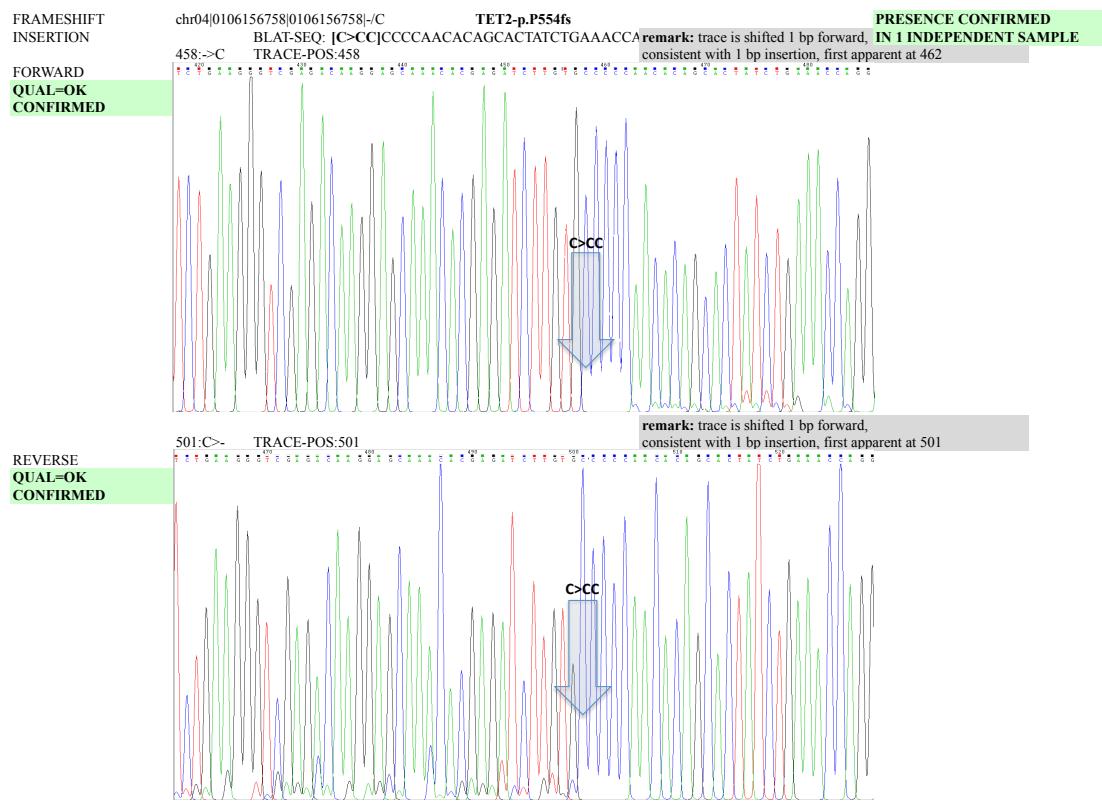


Figure S12: Sanger sequencing results of *TET2* P554fs

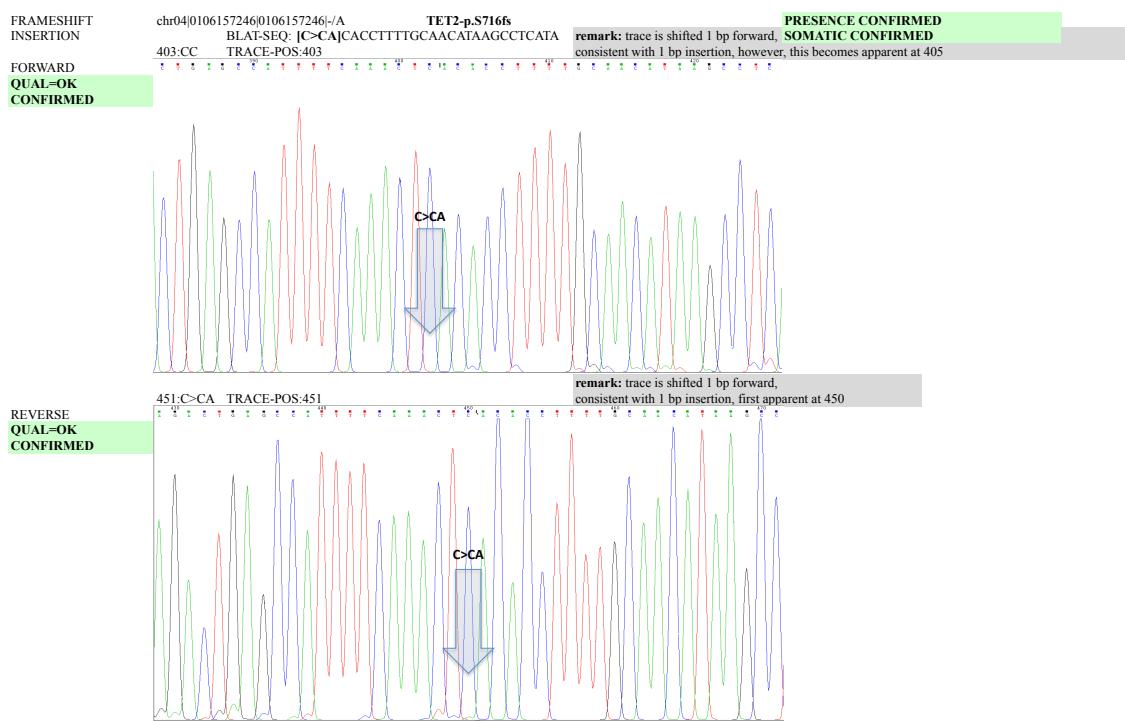


Figure S13: Sanger sequencing results of *TET2* S716fs

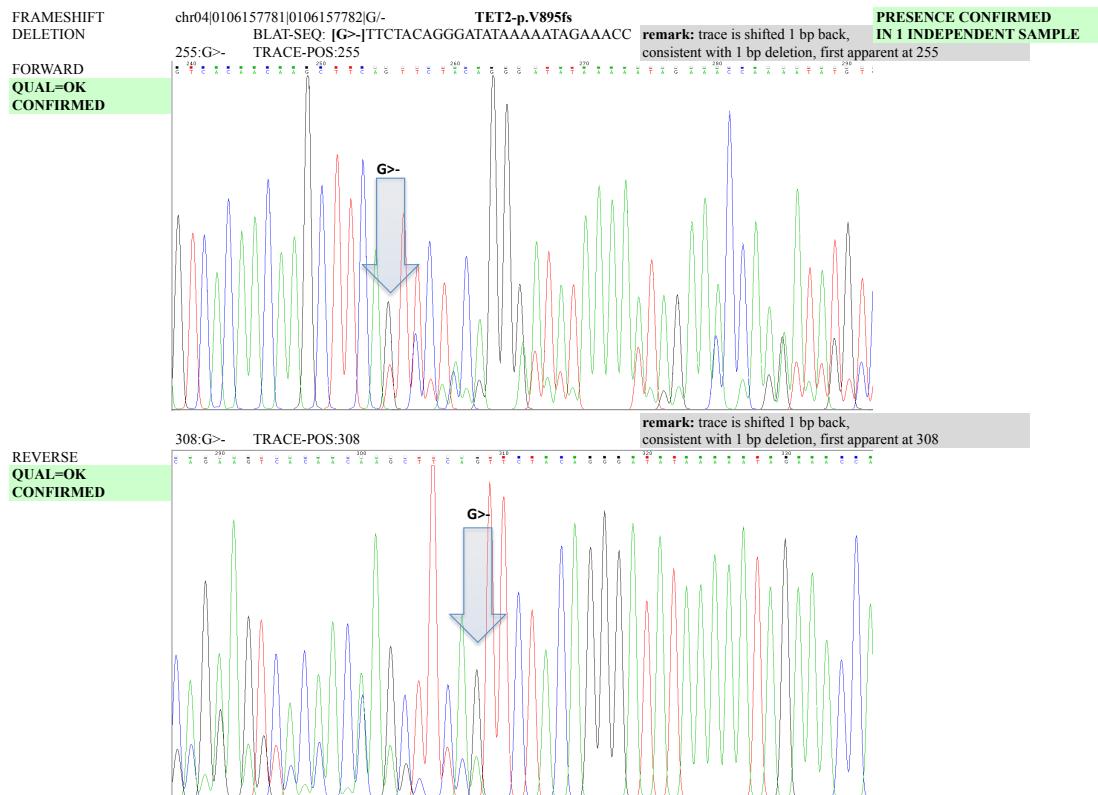


Figure S14: Sanger sequencing results of TET2 V895fs

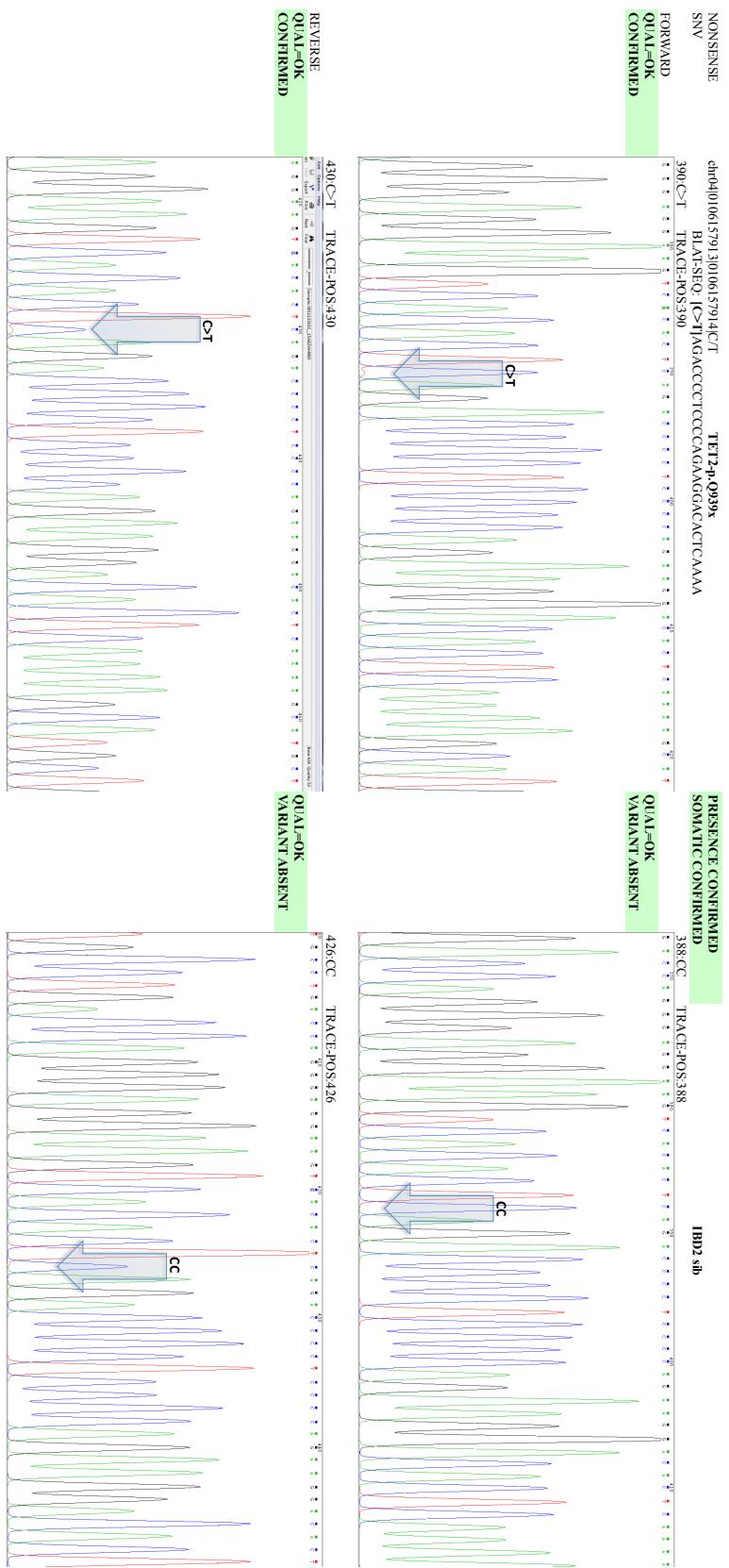


Figure S15: Sanger sequencing results of *TET2* Q939*

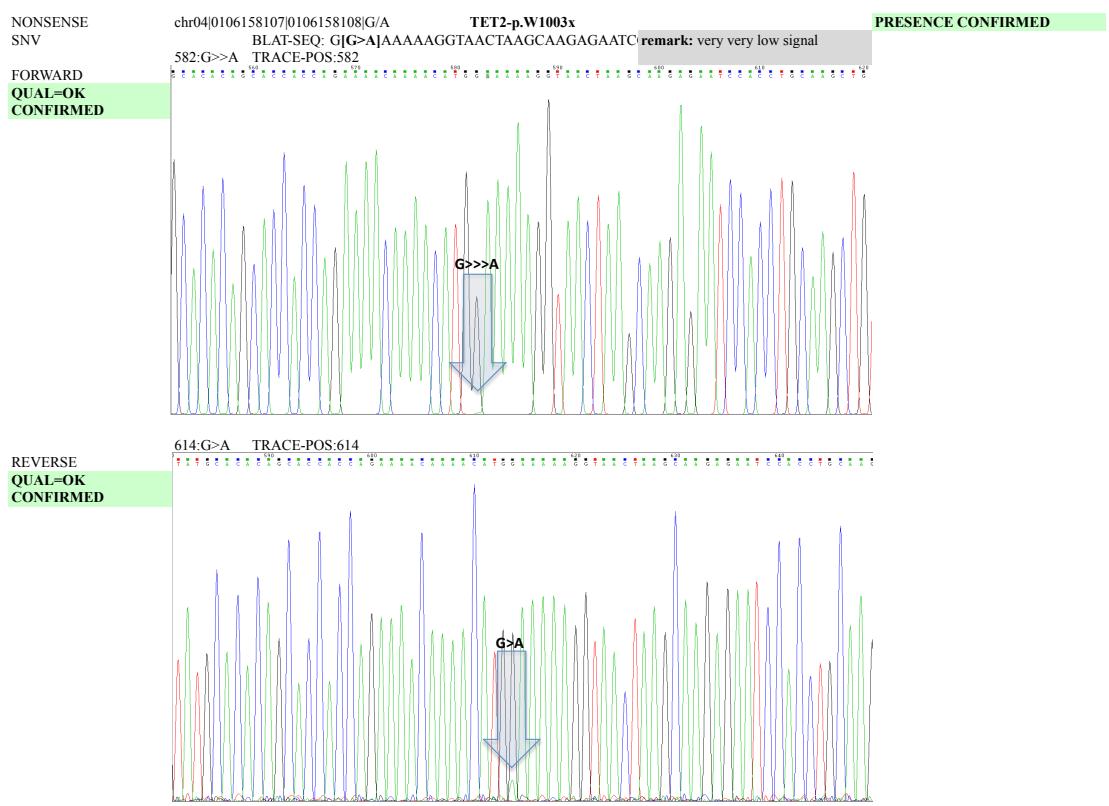


Figure S16: Sanger sequencing results of *TET2* W1003*

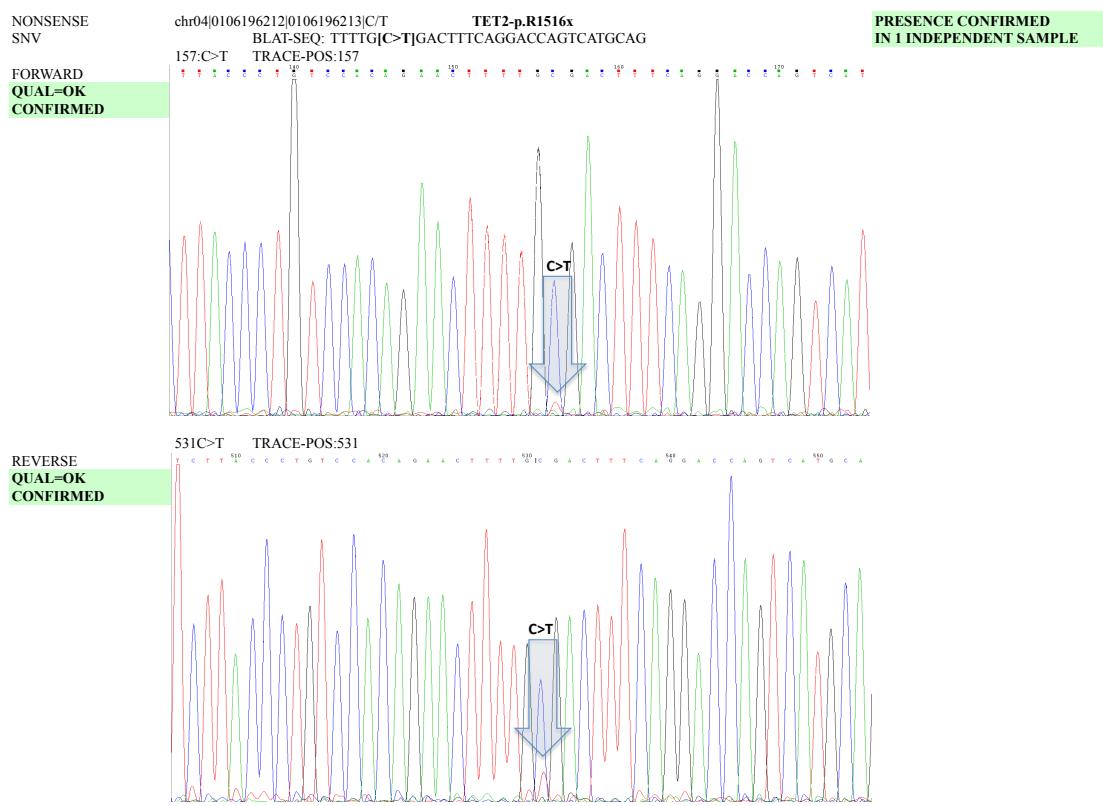


Figure S17: Sanger sequencing results of *TET2* R1516*

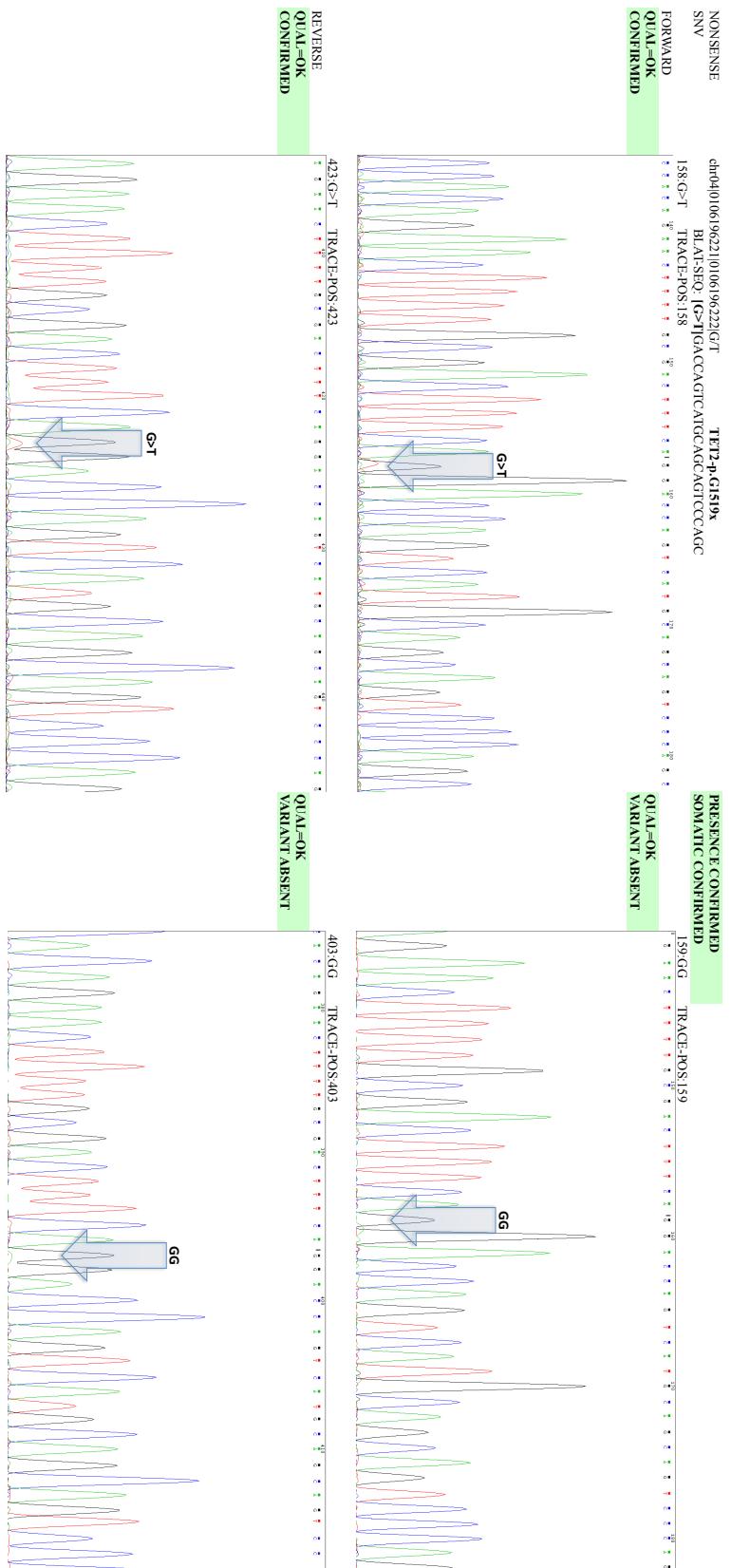


Figure S18: Sanger sequencing results of *TET2* G1519*

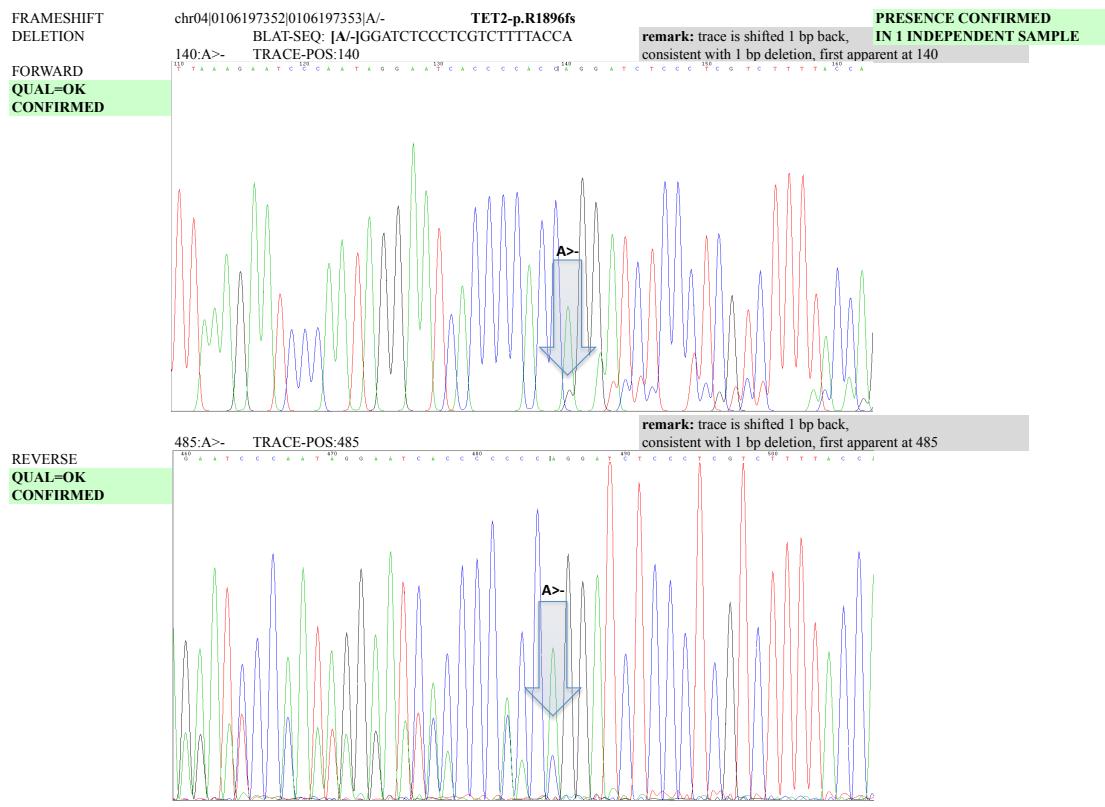


Figure S19: Sanger sequencing results of *TET2* R1896fs

REFERENCES

1. Hofman A, van Duijn CM, Franco OH, et al. The Rotterdam Study: 2012 objectives and design update. *Eur. J. Epidemiol.* 2011;26(8):657–86.
2. Schoenmaker M, de Craen AJM, de Meijer PHEM, et al. Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur. J. Hum. Genet.* 2006;14(1):79–84.
3. Boomsma DI, Wijmenga C, Slagboom EP, et al. The Genome of the Netherlands: design, and project goals. *Eur. J. Hum. Genet.* 2014;22(2):221–7.
4. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60.
5. McKenna N, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297–303.
6. Jun G, Flickinger M, Hetrick KN, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am. J. Hum. Genet.* 2012;91(5):839–848.
7. Genovese G, Kähler AK, Handsaker RE, et al. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *N. Engl. J. Med.* 2014;141126135947008.
8. Jaiswal S, Fontanillas P, Flannick J, et al. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N. Engl. J. Med.* 2014;1–11.
9. Lawrence M, Huber W, Pagès H, et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 2013;9(8):e1003118.
10. R Core Team. R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput. Vienna Austria.* 2014;0:{ISBN} 3–900051–07–0.
11. Xie M, Lu C, Wang J, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* 2014;(Cll.):
12. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* 2014;43(Database issue):D805–11.
13. Li H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics.* 2014;1–9.

14. Genovese G, Handsaker RE, Li H, Kenny EE, McCarroll SA. Mapping the human reference genome's missing sequence by three-way admixture in Latino genomes. *Am. J. Hum. Genet.* 2013;93(3):411–21.
15. Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE. Segmental Duplications: Organization and Impact Within the Current Human Genome Project Assembly. *Genome Res.* 2001;11(6):1005–1017.
16. Bailey J a, Gu Z, Clark R a, et al. Recent segmental duplications in the human genome. *Science (80-.).* 2002;297(5583):1003–1007.
17. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491(7422):56–65.
18. EVS. Exome Variant Server. *NHLBI GO Exome Seq. Proj.* 2014;
19. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* 2002;30(1):97–101.
20. Deelen J, Beekman M, Uh HW, et al. Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell.* 2011;10(4):686–698.
21. Therneau T, Grambsch P. Modeling Survival Data: Extending the Cox Model. *Technometrics.* 2002;44(1):85–86.
22. Langemeijer SMC, Kuiper RP, Berends M, et al. Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat. Genet.* 2009;41(7):838–42.