Somatic mutations reveal clonal cell populations in atherosclerotic plaques

Steffensen LB[1,2], Dembic M[3,4,5], Jensen PS[2,6], Andersen LvB[3], Lindholt JS[2,7], Houlind KC[8], Thomassen M[3], *Rasmussen LM[2,6]

1. Dept. of Molecular Medicine, University of Southern Denmark, Odense, Denmark
2. Centre for Individualized Medicine in Arterial Diseases (CIMA), Odense University Hospital, Odense, Denmark
3. Dept. of Clinical Genetics, Odense University Hospital, Odense, Denmark
4. Dept. of Mathematics and Computer Science (IMADA), University of Southern Denmark, Odense, Denmark
5. Dept. of Clinical Research, University of Southern Denmark, Odense, Denmark
6. Dept. of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark
7. Dept. of Cardiothoracic Surgery, Odense University Hospital, Odense, Denmark
8. Dept. of Regional Health Research, Lillebælt Hospital, Kolding, Denmark

***Corresponding author:**
Lars Melholt Rasmussen
J. B. Winsløws Vej 4, 5000 Odense C, Denmark
Phone: 0045-65414506
Mail: lars.melholt.rasmussen@rsyd.dk

Abstract

In recent years, the potential involvement of clonal cell populations in atherosclerosis has gained renewed interest. Evidence from independent research groups unambiguously showed that smooth muscle cells clonally expand in experimental atherosclerosis, and clonal hematopoiesis of indeterminate potential (CHIP) was identified as a novel independent risk factor for atherosclerotic cardiovascular disease. However, the extend of clonal population partaking in human atherosclerosis remains elusive.

In this study, whole-exome sequencing of 14 carotid plaques unveiled a landscape of somatic mutations. Based on variant allele frequencies, we found that individual locally expanded clones contributed with up to 15% of the plaque cell population, and in seven patients carrying CHIP, haemopoietic clones were represented in plaques in some cases comprising over 30% of plaque cells.

Taken together, we provide evidence that somatic mutations and clonal cell populations (expanded locally or invading from the circulation) are inherent features of atherosclerosis.

Main text

Atherosclerosis is the focal accumulation of lipid, fibrous tissue, and cells in the intimal layer of arteries. The disease develops over decades and may suddenly manifest as myocardial infarction or ischemic stroke - leading causes of death in the world[1].

The cellular component of atherosclerotic plaques mainly originates from circulating myeloid cells and smooth muscle cells (SMCs) recruited from the local media[2]. Recent advances in the field have revised our perception of both sources: First, by use of multicolor lineage tracing in mouse models, it was unambiguously demonstrated that all SMC-derived plaque cells are progeny of few medial SMCs[3,4], thus reviving the hypothesis of monoclonality in atherosclerosis proposed by Benditt and Benditt almost 50 years ago[5].

Second, several lines of evidence now support a causal role for clonal hematopoiesis of indeterminate potential (CHIP) in atherosclerotic cardiovascular disease[6,7]. Individuals with CHIP carry expanded hematopoietic clones often harboring mutations in epigenetic regulators (*e.g.*, *DNMT3A* and *TET2*)[8], which besides conferring a selective advantage render myeloid cells more inflammatory and pro-atherogenic[9]. Whether similar mechanisms (*i.e.,* somatic mutation-driven clonal expansion) operate locally in plaques is unknown.

Although these recent advances point to the involvement of clonal cell populations (either expanded locally in plaques or invading from the circulation) in the pathology of atherosclerosis, the extend of clonal population partaking in human atherosclerosis remains elusive.

In this study, we performed whole-exome sequencing of DNA extracted from carotid plaques and patient-matched buffy coats from 14 patients undergoing carotid endatherectomy (**supplementary table S1**). From each carotid plaque, one to four segments (2-5 mm tissue per segment) distanced 4-10 mm were analyzed in parallel (**fig. 1a**, **supplementary fig. S1**).

We identified on average 26 (range: 2-83) somatic mutations per plaque (variant allele frequencies (VAF) ≥ 0.01) (**fig. 1b**), which were non-detectable in patient-matched buffy coats. VAFs ranged from 0.01 to 0.096 (median: 0.022) (**supplementary fig. S2a**), representing locally expanded clones constituting 1.4-15.4% (median: 4.3%) of total plaque cell populations (**fig. 1c-d**, **supplementary fig. S2b**). Somatic mutations were detected in 29 of the 32 sequenced plaque segments, and seven clones spanned several segments of a plaque requiring a physical extend of individual clones of at least 4-16 mm (**fig. 1e**). To evaluate the number of cells constituting individual clones, we calculated the number of sequenced cells from plaque segment DNA yields, assuming 6 pg DNA per cell. Cell numbers ranged from 77,000 to 1,278,000 (median: 605,000) resulting in clone cell number estimates of 1,483-112,327 (median: 22,756). However, since DNA purifications may be sub-efficient, these numbers are presumably underestimated.

After establishing the existence of locally expanded clones, we sought to evaluate whether somatic mutations were clone *drivers*, *i.e.*, conferring a selective advantage (analogous to tumors), or merely amplified by coincidence as *passengers* in a clone expanding by other mechanisms. The overall ratio of detrimental mutations (*i.e.*, loss-of-function and missense mutations) to mutations not affecting the protein residue sequence (*i.e.*, synonymous mutations or mutations located outside coding sequences) was 2.6 (263/103) – an indication that somatic mutations do not occur by random, which is compatible with some mutations possessing clone-driving properties[10] (**fig. 1b**).

We did not detect any mutational hotspots (as in CHIP[8] and certain types of cancer[11]), *i.e.,* genes mutated in several plaques. However, 17 of the mutated genes appeared in a recently reported compendium of 568 mutational cancer driver genes[11] (χ^2 test: $p < 0.001$).

To further investigate potential functional enrichment of mutated genes, we used two approaches. First, over-representation analysis of all 346 mutated genes showed enrichment of "metal ion transmembrane transporter"-, "actin binding"-, "voltage-gated channel"-, and "protein tyrosine kinase"-activity terms (**fig. 1f**, **supplementary fig. S3a-b**). Tyrosine kinase activity typically mediates intracellular signaling associated with cell growth and motility implying potential clonal driver effects of these mutations, and mutated tyrosine kinases are frequently implicated in tumor development[12].

Second, as a complementary approach to assess functional enrichment, we next used gene expression data from the gene-tissue expression (GTEx) portal[13] to compare distribution of gene expression levels of the 346 mutated genes with that of all genes of the genome (**fig. 1g**), and found the distribution to be skewed toward higher gene expression (χ^2 test: $p < 0.00001$). This was not specific for artery tissue as the same was observed in other tissues, such as liver and spleen (**supplementary fig. S3a-e**).

Taken together, although we did not identify hotspot clonal driver genes, three independent lines of evidence (enrichment of detrimental mutations, functional enrichment of mutated genes, and increased gene activity of mutated genes) were compatible with a non-random distribution of mutations throughout the exome, suggesting contribution of somatic mutations to clonal expansion. These observations add atherosclerosis to the growing number of non-cancerous diseases in which somatic mutations are present and potentially participate in disease development[14].

To gain more insight into mutational processes that could underlie the accumulation of somatic mutations in atherosclerotic plaques, we next investigated enrichment of mutational signatures. Cytosine to thymine transitions (C>G) accounted for >40% of mutations (**supplementary fig. S3a**) and most frequently occurred next to a 3' guanine (**supplementary fig. S3b**) indicating methylation and de-amination as a possible mutational mechanism. By analysis of mutational signatures using the Catalogue of Somatic Mutations in Cancer (COSMIC) database[15], we found enrichment of single base substitution (SBS) signatures previously associated with "thiopurine chemotherapy treatment" (SBS87), "tobacco smoking" (SBS5), "defective DNA mismatch repair" (SBS6) and "ultraviolet radiation" (SBS7a/b) (**supplementary fig. S3c-e**). Some of these terms are likely not relevant for a tissue that has not been exposed to chemotherapy or ultraviolet radiation, however, mutational signatures are annotated by association in the context of cancers, and future studies may unveil whether plaque-specific mutational processes result in similar signatures.

We next investigated the contribution of circulating CHIP mutation-carrying cells to the plaque cell population. By screening for mutations in 78 previously established CHIP genes[8] using buffy coat DNA, we found 7 of the 14 patients to be CHIP carriers, and several CHIP mutations were detected in each of these patients (**fig. 2a**). Buffy coat derived VAFs ranged from 0.01 to 0.17 (median: 0.05) (**fig. 2a**) corresponding to clone cell frequencies of 2-34% (median: 10%). Except for a minor *DNMT3A* clone in patient 13, all CHIP clones were represented in patient-matched carotid plaques and were evenly distributed across plaque segments (**fig. 2b**). This distribution contrasted that of locally expanded clones, which was more confined to individual plaque segments. Indeed, this would be expected from a more diffuse clonal population invading from the circulation as compared to more immobilized focally expanding clones within the plaque. CHIP clone fractions in buffy coats correlated with plaque fractions of the corresponding clones, and CHIP clones entering from the circulation contributed with as much as 42% of plaque cells (**fig. 2c**).

The present study provides evidence that the growth of atherosclerotic plaques does not result from simultaneous division of all cells at an equal rate. Rather, our findings are compatible with individual cells proliferating to form clones, which make up considerable portions of human carotid plaque cells. Although our data indicate that the mutational spectrum is not random, it remains elusive to what extend somatic mutations contribute to clonal expansion. Indeed, as proposed for clonal expansion in mouse models[4,16], clone precursors may be a subset of cells possessing advantageous properties related to *e.g.*, proliferation, survival in a toxic plaque environment, or escaping counter-regulatory mechanisms of the immune system. Alternatively, a clone-founding cell could merely be located "at the right place at the right time" thereby be the first to supersede a stimulatory threshold for activation, and as previously proposed, the selected clone may simultaneously suppress expansion of neighboring cells through juxtacrine lateral inhibition[4]. In this scenario, randomly amplified passenger mutations detected by whole-exome sequencing would only represent a fraction of clones (**supplementary fig. S7c-d**). Finally, clonal patches carrying somatic mutations could already be present in the carotid artery prior to plaque development (**supplementary fig. S7e-f**). Indeed, some degree of clonality has been observed in most normal-appearing tissues, including blood vessels[5,17,18].

A limitation to the analytical approach employed in this study is the inability to assess distribution of mutations across clonal populations within each plaque. Few clones could harbor numerous mutations, and/or multiple clones could harbor few mutations (**supplementary fig. S3a-b**). Indeed, for six of the patients, the sum of clone cell frequencies estimated from each mutation exceeds 100%, excluding a scenario of a single mutation per clone in these patients. Moreover, our data do not allow us to pinpoint which cell type(s) clones are derived from or how clones are physically organized within plaque tissue.

In recent decades, the epidemiology of atherosclerotic cardiovascular disease has changed from being a disease of middle-aged smoking, hypercholesterolemic, hypertensive men, to increasingly being a disease of older people[19]. This epidemiological transition is attributed to the successes of lipid- and blood pressure-lowering medication and smoking cessation initiatives, as well as a growing ageing population in most countries[19]. As a consequence, the residual risk is increasingly attributed to age-related unknowns. The identification of CHIP as an age-associated independent risk factor highlights somatic mutations and clonally expanded cells as significant contributors to atherosclerotic cardiovascular disease[6,7]. The present study demonstrates that similar processes take place locally in plaque tissues, but to what extent local somatic mutations and clonal expansion contribute to disease development and residual risk remains to be investigated. To this end, another important notion from our study is that somatic mutations may serve as genetic tags for identification and characterization of clonal populations in future studies. This may prove to be a convenient tool for deciphering clonal biology in atherosclerosis to obtain unprecedented insight that could have implications for future therapeutics.

Methods

Ethical approval and patient sample processing

The study was approved by the Danish National Committee on Health Research Ethics (Project ID: CVK-2006749) and The Regional Committees on Health Research Ethics for Southern Denmark (Project ID: S-20140202) and is registered in the Registry of Research Projects in the Region of Southern Denmark (file no. 18/20280).

Carotid plaque tissue and buffy coats from 14 later deceased patients who had undergone carotid endatherectomy either at the Department of Cardiac, Thoracic and Vascular Surgery, Odense University Hospital (OUH) or the Department of Vascular Surgery, Kolding Hospital were included in the study. Patients were admitted for surgery because of carotid plaque-associated symptoms. Informed consent was obtained from all participants prior to surgery.

Immediately after surgery, carotid plaque samples were placed and kept in ice cold Hanks Buffered Salt Solution containing 10 mM HEPES (Biological Industries, SKU BI-02-016-1A and BI-03-025-1B, respectively). Within 24 hours, plaques were sliced into 2 (patients 2, 4, 10, 13, and 14) or 5 mm thick segments (patients 1, 3, 5, 6, 7, 8, 9, 11, and 12) using a custom-made device. Every third segment was frozen at -80°C and stored in the Odense Artery Biobank infrastructure of the Centre for Individualized Medicine in Arterial diseases (CIMA) at OUH.

Blood was collected in EDTA vials, centrifuged at 4,000 x *g* for 10 minutes at 4°C, and buffy coats were collected using a pipette from the top of the centrifuged sample, and stored at -80°C in the Odense Artery Biobank.

Whole-exome sequencing and data analysis

Whole-exome sequencing was performed on a total of 32 plaque segments and 14 buffy coat samples from the 14 patients. Genomic DNA was isolated using standard methods. Libraries were prepared using Twist Comprehensive Exome Panel kit (Twist Bioscience, San Francisco, US), and targeted 2x150 bp pair-end sequencing was performed on an Illumina NovaSeq 6000 platform.

Fastq files were aligned to reference sequence (human NCBI build 38) using the Burrows-Wheeler algorithm (BWA-MEM) v.0.7.12 and processed with the Genome Analysis Toolkit (GATK 4.2.0.0, gatk.broadinstitute.org) to generate recalibrated BAM files. For the identification of single nucleotide variants and indels occurring in both buffy coat and plaque tissue, we used MuTect2 in single sample mode. To identify somatic variants occurring only in the plaque samples, we performed joint variant-calling on matched samples plaque vs. buffy coat. Variants were annotated using VarSeq version 2.2.1 (Golden Helix, Inc.) and filtered through a cascade of filters. Briefly, for variant calling in non-matched samples, the filters were: (i) variant is in a CHIP gene according to a recently published list of 78 CHIP genes[8], (ii) variant has more than 20 reads, (iii) variant has a known or predicted damaging effect on protein functionality (according to ClinVar or prediction tools), and (iv) variant has a minor allele frequency (MAF) < 0.01 (according to gnomAD). The Genome Aggregation Database (gnomAD) is the world largest public collection of genetic variation, and it is based on 125,748 exomes and 15,708 whole-genomes collected worldwide (https://gnomad.broadinstitute.org/). We used the gnomAD collection from exome and genome sequencing data as a large external control group. For matched MuTect2 calling plaque vs. buffy coat, we selected variants with (i) >5 reads, (ii) a variant allele frequency (VAF) in reads > 1%, and no or very low signal in the buffy coat (< 0.01%). Built-in post-calling filters in MuTect2 were used to select for high quality variants. All filtered variants were inspected manually using VarSeq.

For functional prediction of nonsynonymous variants, we used the Combined Annotation Dependent Depletion (CADD Scores 1.4) method (PHRED scaled scores)[20]. The CADD algorithm includes conservation metrics, functional genomic data, exon-intron boundaries, and protein functionality scores. A score of ≥20 indicates that a variant is predicted to be among the 1% of the most deleterious substitutions. We used a CADD score of >23 to select for a damaging effect. In addition, we used a dbNSFP tool (database for nonsynonymous single nucleotide polymorphisms' functional predictions) which incorporates five different algorithms (SIFT, Polyphen2, MutationTaster, MutationAssessor, and FATHMM)[21]. The most damaging variants were selected based on a voting system where 4 or all 5 of these algorithms predicted a negative functional impact. Data are available as **supplementary table S2-S4**.

Mutational signature analysis

Mutational signatures analysis was performed using the MutationalPatterns R packages (v3.6.0)[22]. Due to the limited number of mutations across all samples (n = 367), mutations were pooled to increase analytical power. A substitution catalogue was constructed from the frequency of the possible 96 trinucleotides. The low number of mutations did not enable us to extract any novel signatures. Instead, the catalogue of substitutions was fitted to the known COSMIC reference signatures (v3.2)[23].

To minimize the risk of overfitting and misattribution, the stricter refitting approach of MutationalPatterns was used. This approach starts with a standard fitting on all the COSMIC reference signatures and then in an iterative fashion removes the signature with the lowest contribution. In each iteration, the cosine similarity between the original and reconstructed

profile is calculated. The iterations stop when the difference between two iterations becomes larger than a certain cutoff. To verify the stability of the resulting mutational signatures, the strict refitting was performed in 10,000 iterations using bootstrapping. We tested multiple cutoffs and found that a cutoff around 0.01 provided results that were stable through bootstrapping in the sense that they were included in the refitting solution 50% or more.

Disclosures

The authors have nothing to declare.

Data availability statement

The study was approved by the Danish National Committee on Health Research Ethics (Project ID: CVK-2006749and S-20140202) and is registered in the Registry of Research Projects in the Region of Southern Denmark (file no. 18/20280). Carotid plaque tissue and buffy coats from 14 later deceased patients who had undergone carotid endatherectomy either at the Department of Cardiac, Thoracic and Vascular Surgery, Odense University Hospital (OUH) or the Department of Vascular Surgery, Kolding Hospital were included in the study. Patients were admitted for surgery because of carotid plaque-associated symptoms. Informed consent was obtained from all participants prior to surgery.

Acknowledgements

We thank Mikkel Engstrøm Graversen for discussion of the statistical analyses used in the study.

Funding

The study was supported by a grant from Odense University Hospital to the Center for Individualized medicine in Arterial diseases (CIMA) and a grant from the Novo Nordisk Foundation to LMR (NNF20OC0065744).
LBS is supported by Danish Diabetes Academy (supported by the Novo Nordisk Foundation).

References

1 Diseases, G. B. D. & Injuries, C. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1204-1222, doi:10.1016/S0140-6736(20)30925-9 (2020).

2 Libby, P. *et al.* Atherosclerosis. *Nat Rev Dis Primers* **5**, 56, doi:10.1038/s41572-019-0106-z (2019).

3 Chappell, J. *et al.* Extensive Proliferation of a Subset of Differentiated, yet Plastic, Medial Vascular Smooth Muscle Cells Contributes to Neointimal Formation in Mouse Injury and Atherosclerosis Models. *Circ Res* **119**, 1313-1323, doi:10.1161/CIRCRESAHA.116.309799 (2016).

4 Jacobsen, K. *et al.* Diverse cellular architecture of atherosclerotic plaque derives from clonal expansion of a few medial SMCs. *JCI Insight* **2**, doi:10.1172/jci.insight.95890 (2017).

5 Benditt, E. P. & Benditt, J. M. Evidence for a monoclonal origin of human atherosclerotic plaques. *Proc Natl Acad Sci U S A* **70**, 1753-1756, doi:10.1073/pnas.70.6.1753 (1973).

6 Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* **371**, 2488-2498, doi:10.1056/NEJMoa1408617 (2014).

7 Jaiswal, S. *et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med* **377**, 111-121, doi:10.1056/NEJMoa1701719 (2017).

8 Marnell, C. S., Bick, A. & Natarajan, P. Clonal hematopoiesis of indeterminate potential (CHIP): Linking somatic mutations, hematopoiesis, chronic inflammation and cardiovascular disease. *J Mol Cell Cardiol* **161**, 98-105, doi:10.1016/j.yjmcc.2021.07.004 (2021).

9 Fuster, J. J. *et al.* Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science* **355**, 842-847, doi:10.1126/science.aag1381 (2017).

10 Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**, 1029-1041 e1021, doi:10.1016/j.cell.2017.09.042 (2017).

11 Martinez-Jimenez, F. *et al.* A compendium of mutational cancer driver genes. *Nat Rev Cancer* **20**, 555-572, doi:10.1038/s41568-020-0290-x (2020).

12 Saraon, P. *et al.* Receptor tyrosine kinases and cancer: oncogenic mechanisms and therapeutic approaches. *Oncogene* **40**, 4079-4093, doi:10.1038/s41388-021-01841-2 (2021).

13 Consortium, G. T. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213, doi:10.1038/nature24277 (2017).

14 Mustjoki, S. & Young, N. S. Somatic Mutations in "Benign" Disease. *N Engl J Med* **384**, 2039-2052, doi:10.1056/NEJMra2101920 (2021).

15 Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**, D941-D947, doi:10.1093/nar/gky1015 (2019).

16 Wang, Y. *et al.* Clonally expanding smooth muscle cells promote atherosclerosis by escaping efferocytosis and activating the complement cascade. *Proc Natl Acad Sci U S A* **117**, 15818-15826, doi:10.1073/pnas.2006348117 (2020).

17 Yizhak, K. *et al.* RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science* **364**, doi:10.1126/science.aaw0726 (2019).

18 Murry, C. E., Gipaya, C. T., Bartosek, T., Benditt, E. P. & Schwartz, S. M. Monoclonality of smooth muscle cells in human atherosclerosis. *Am J Pathol* **151**, 697-705 (1997).

19 Libby, P. The changing landscape of atherosclerosis. *Nature* **592**, 524-533, doi:10.1038/s41586-021-03392-8 (2021).

20 Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886-D894, doi:10.1093/nar/gky1016 (2019).

21 Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-315, doi:10.1038/ng.2892 (2014).

22 Manders, F. *et al.* MutationalPatterns: the one stop shop for the analysis of mutational processes. *BMC Genomics* **23**, 134, doi:10.1186/s12864-022-08357-3 (2022).

23 Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94-101, doi:10.1038/s41586-020-1943-3 (2020).

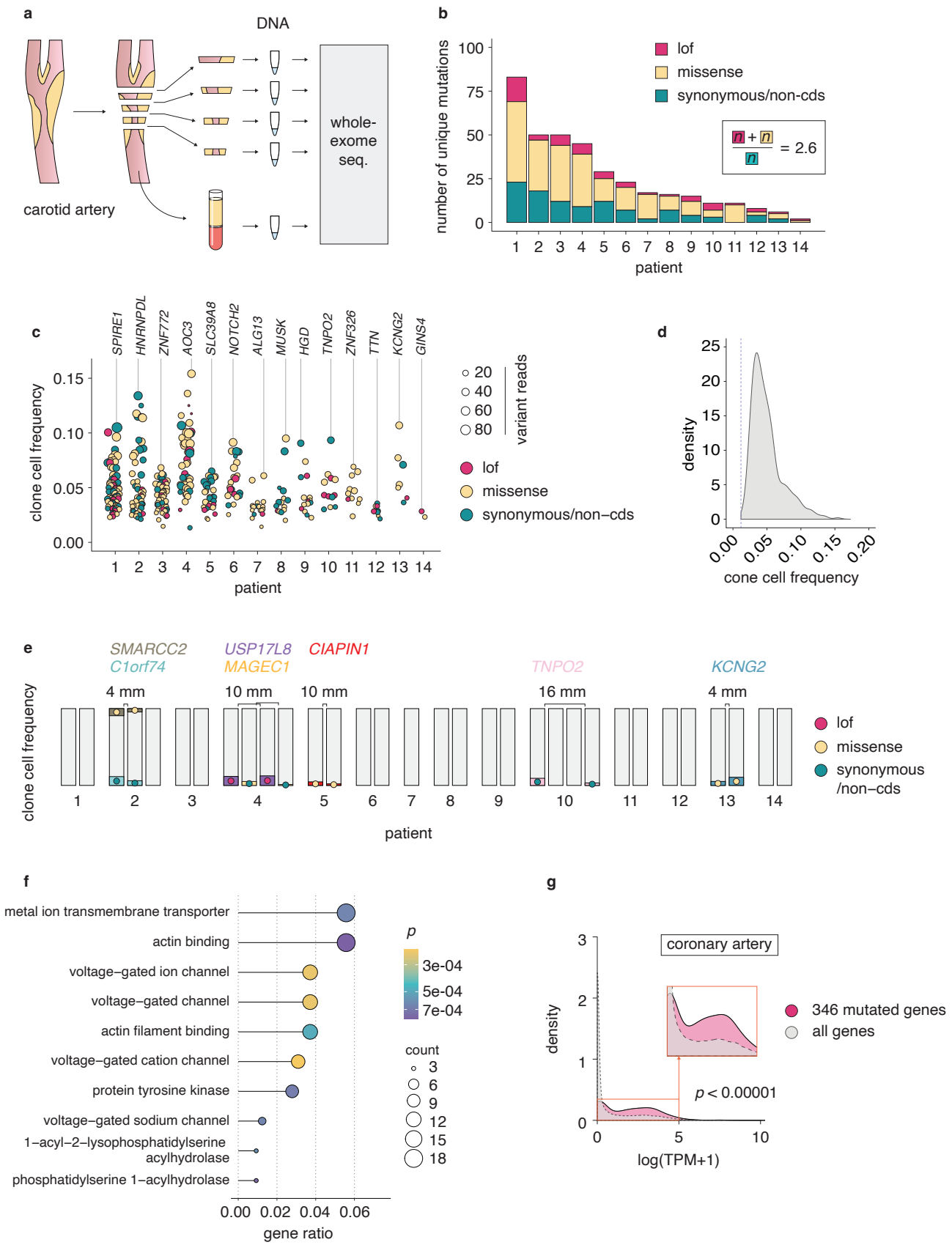

a
DNA
carotid artery
whole-exome seq.
b
number of unique mutations
lof
missense
synonymous/non-cds
= 2.6
patient
c
SPIRE1
HNRNPDL
ZNF772
AOC3
SLC39A8
NOTCH2
ALG13
MUSK
HGD
TNPO2
ZNF326
TTN
KCNG2
GINS4
clone cell frequency
variant reads
lof
missense
synonymous/non–cds
patient
d
density
cone cell frequency
e
SMARCC2
C1orf74
USP17L8
MAGEC1
CIAPIN1
TNPO2
KCNG2
4 mm
10 mm
10 mm
16 mm
4 mm
clone cell frequency
lof
missense
synonymous /non–cds
patient
f
metal ion transmembrane transporter
actin binding
voltage–gated ion channel
voltage–gated channel
actin filament binding
voltage–gated cation channel
protein tyrosine kinase
voltage–gated sodium channel
1–acyl–2–lysophosphatidylserine acylhydrolase
phosphatidylserine 1–acylhydrolase
gene ratio
p
count
g
coronary artery
density
346 mutated genes
all genes
p < 0.00001
log(TPM+1)

Figure 1 | Locally expanded clone populations.

a. Carotid atherosclerotic plaques (one to four segments per plaque) and buffy coats from 14 patients undergoing carotid endatherectomy were analyzed by whole-exome sequencing. **b.** Barplot showing the number of plaque-confined somatic mutations (*i.e.*, mutations not identified in patient-matched buffy coats) for each patient. Mutation type is indicated by color. lof = loss of function; cds = coding sequence. **c.** Frequency of clone cells carrying a specific somatic mutation in carotid plaques. Clone cell frequencies are calculated based on the assumption of mutation heterozygosity by multiplying variant allele frequencies (VAFs) (shown in supplementary fig. S2a) by two, except for mutations located on the X-chromosome in male patients. Although single clones may harbor several mutations, each mutation is represented by separate clone cell frequencies on the plot. For each patient, the identity of the mutated gene representing the highest clone cell frequency is shown. Mutation type is indicated by dot color, and variant reads is indicated by dot size. **d.** The density plot shows the distribution of clone cell frequencies calculated from each somatic mutation VAF from all patients (similar plots are shown for individual patients in supplementary fig. S2b). The dashed line shows the defined limit of detection. **e.** Seven somatic mutations spanned more than one plaque segment. The total cell population within each plaque segment is represented by a grey bar, and the colored areas correspond to the fraction of cells harboring one of the seven mutations. Mutation type is indicated by dot color. Intersegment distances used to estimate physical extends of clonal populations are indicated. **f.** Dotplot showing the ten most enriched molecular function gene ontology terms (using clusterProfiler 4.0 R package[20]) of the 346 genes mutated across all patients. Dot sizes indicate number of mutated genes for each term. The ratio shows the coverage of a given term by mutated genes, and dot colors indicate significance levels. **g.** Density plot showing the distribution of gene expression levels (presented as the logarithm of transcripts per million (TPM) +1) in coronary artery tissue (data from the gene-tissue expression (GTEx) portal, v8) for all genes of the genome and for the 346 mutated genes identified in carotid plaques. Area under the curves have been adjusted to be equal for the two datasets. To test if distributions were significantly different, a χ^2 test was performed based on the number of genes with expression levels above and below the median expression level of mutated genes, respectively.

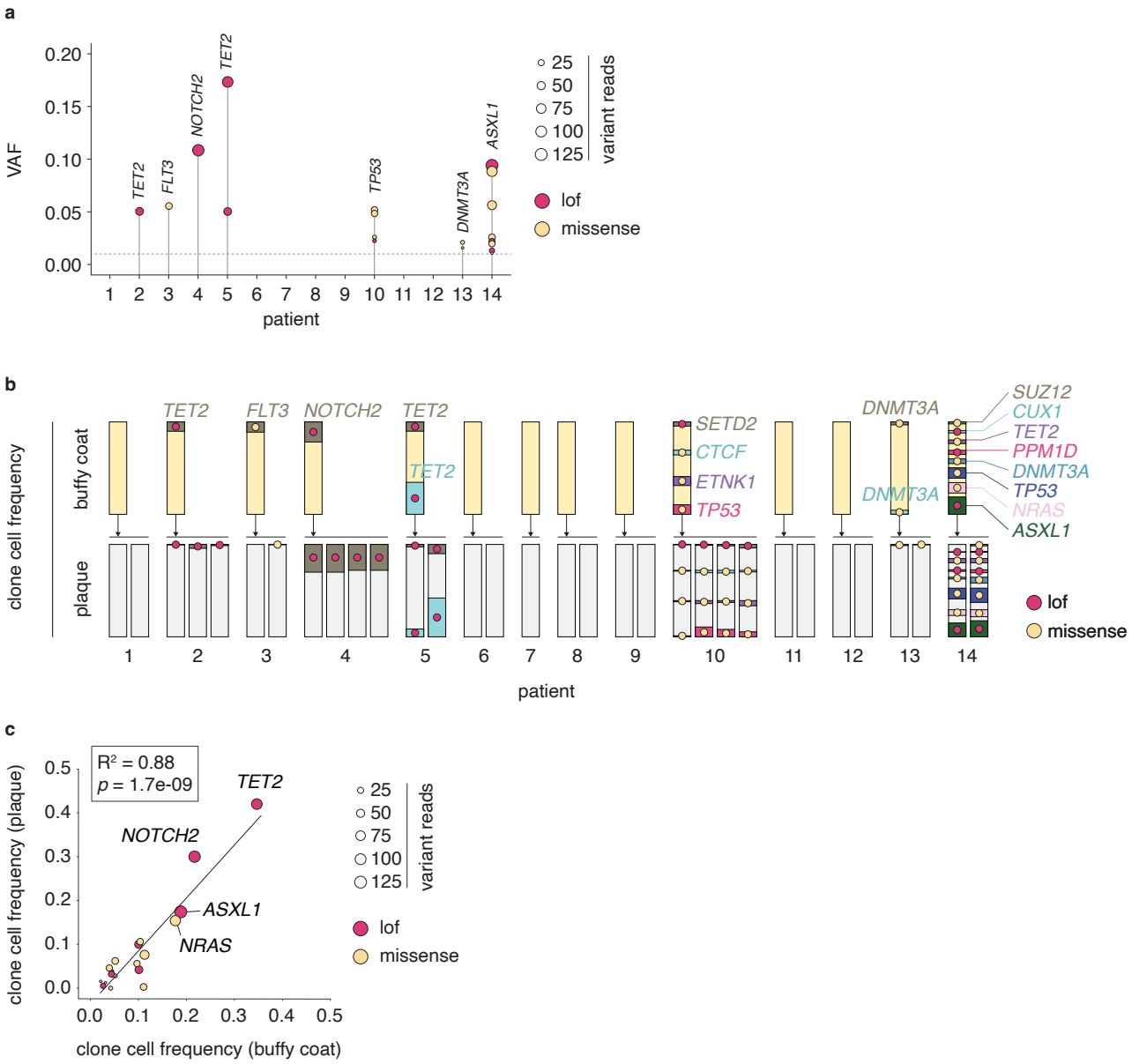
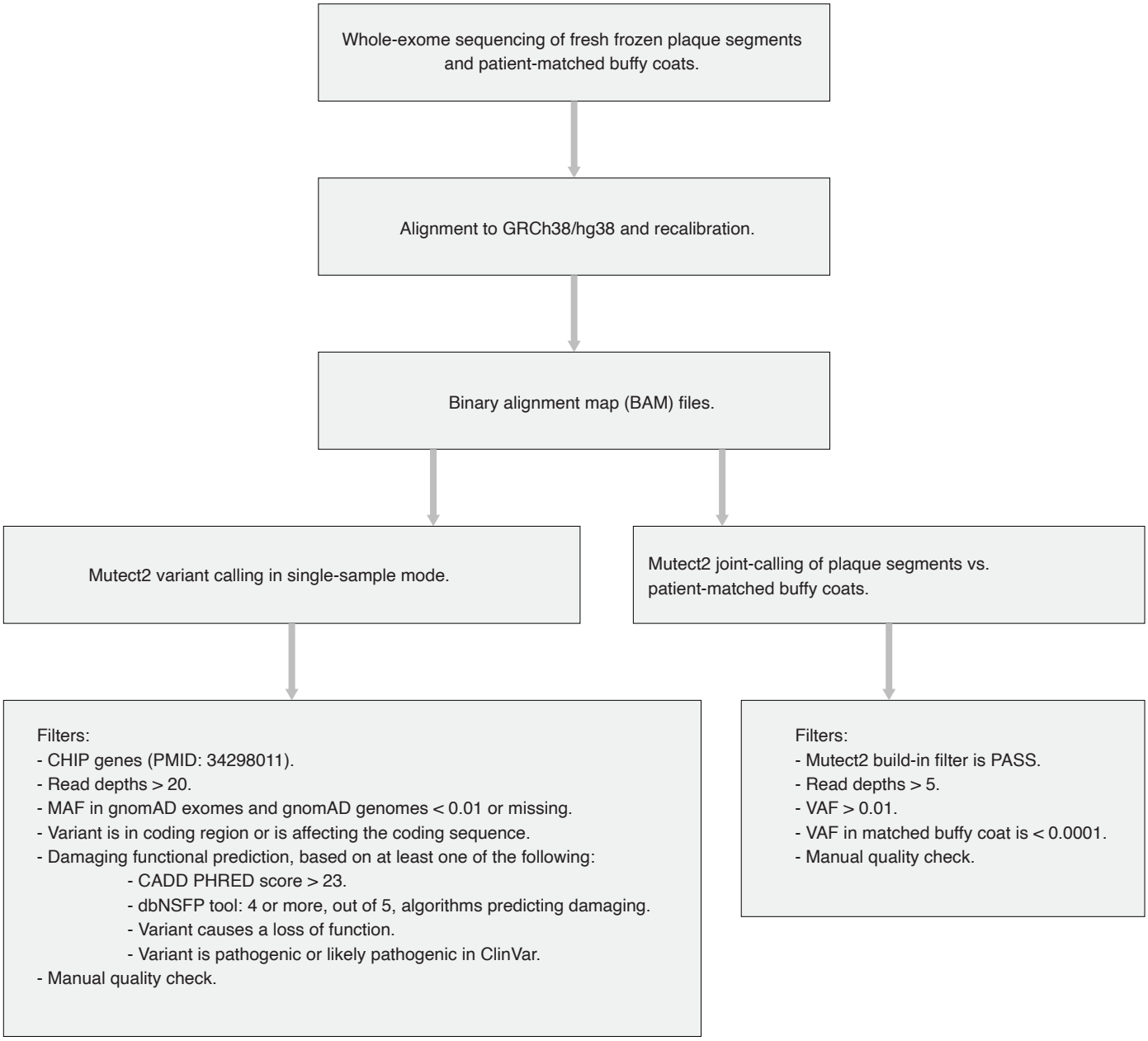
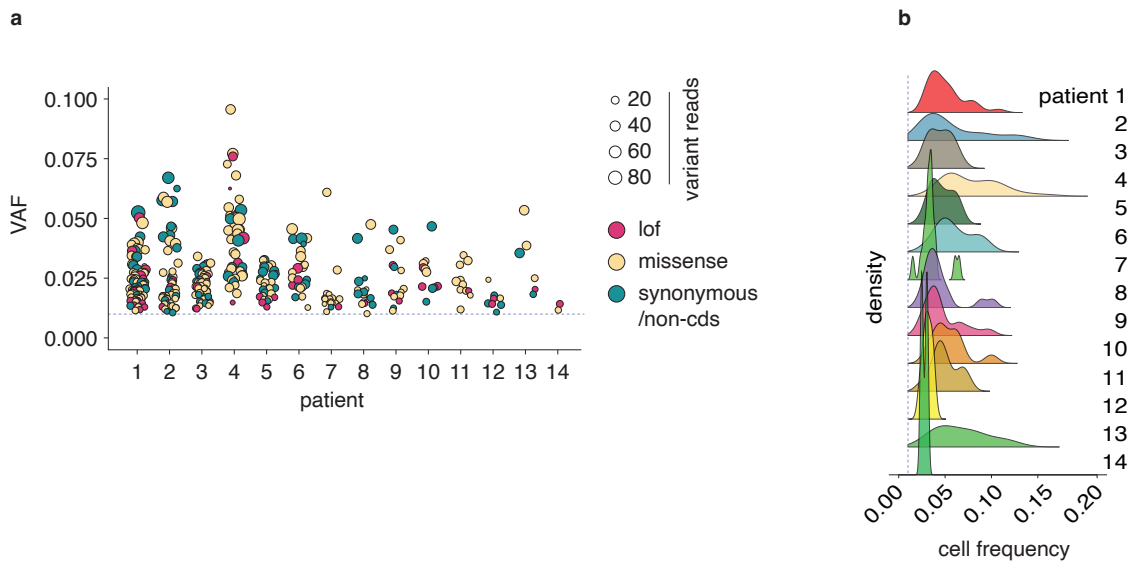


Figure 2 I Contribution of clonal hematopoiesis of indeterminate potential (CHIP) clones to the carotid plaque cell population.

a. Seven of the 14 patients were CHIP-carriers. The plot shows variant allele frequencies (VAFs) of CHIP mutations detected in buffy coats. For each patient, the identity of the mutated gene with the highest VAF is shown. Dot sizes indicate variant reads, and colors indicate the type of mutation. lof = loss of function. The dashed line shows the defined limit of detection. **b.** CHIP clone cell frequencies in buffy coats (yellow bars) and in plaque segments (grey bars) from the 14 patients. Clone cell frequencies are represented by colored areas within yellow and grey bars. Mutation type is indicated by dot color. **c.** Correlation of the frequency of CHIP cells carrying a particular somatic mutation in buffy coats and the maximal clone cell frequency across segments from patient-matched carotid plaques. CHIP gene identities are shown for the four clones with highest cell frequencies. Dot sizes indicate variant reads, and colors indicate the type of mutation. Statistics are based on *Pearson* correlation coefficient.

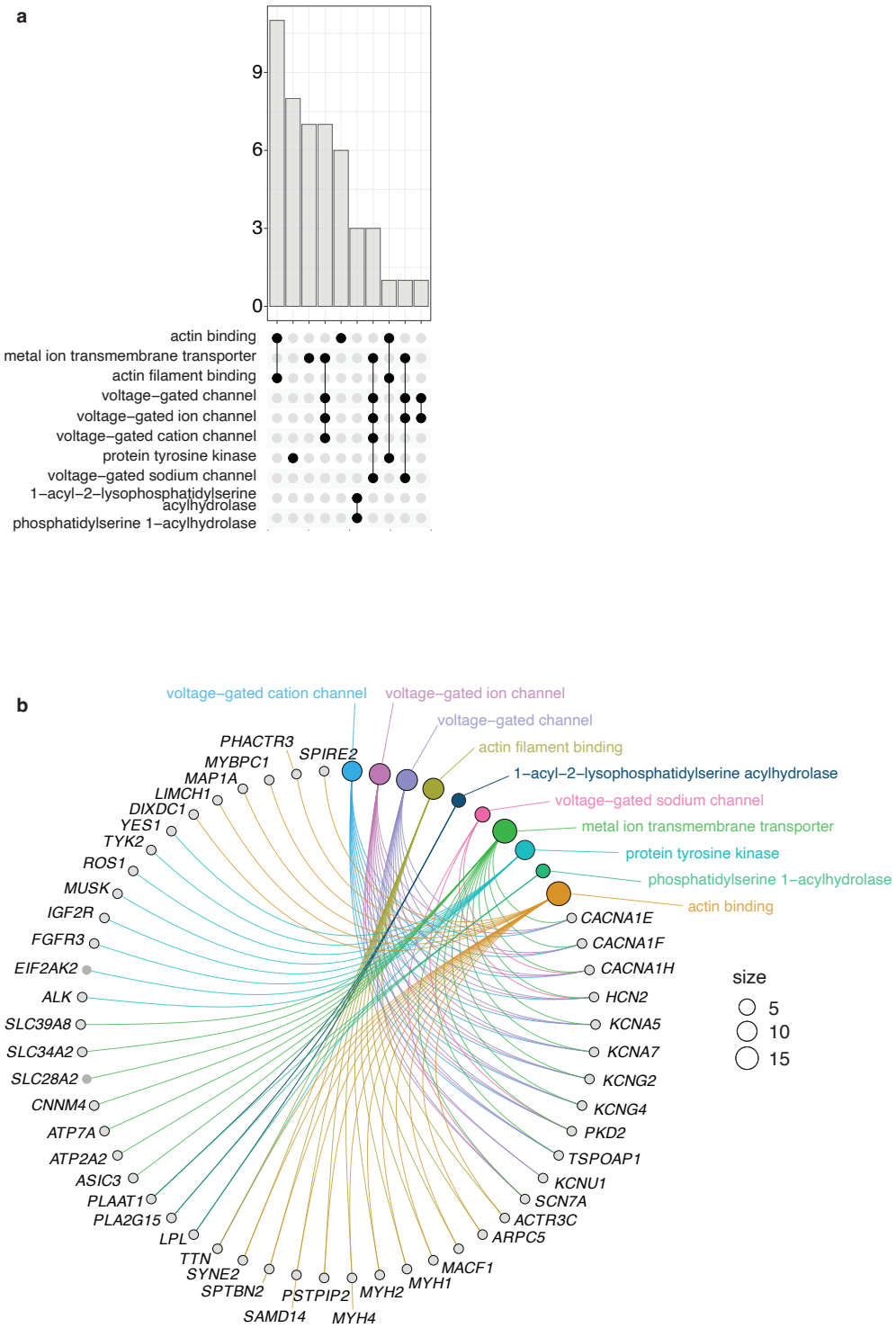


Supplementary figure S1 I
Flowchart showing the genetic analysis of the study. MAF (minor allele frequency); gnomAD (Genome Aggregation Database); CADD PHRED (Combined Annotation Dependent Depletion PHRED scale); VAF (variant allele frequency).



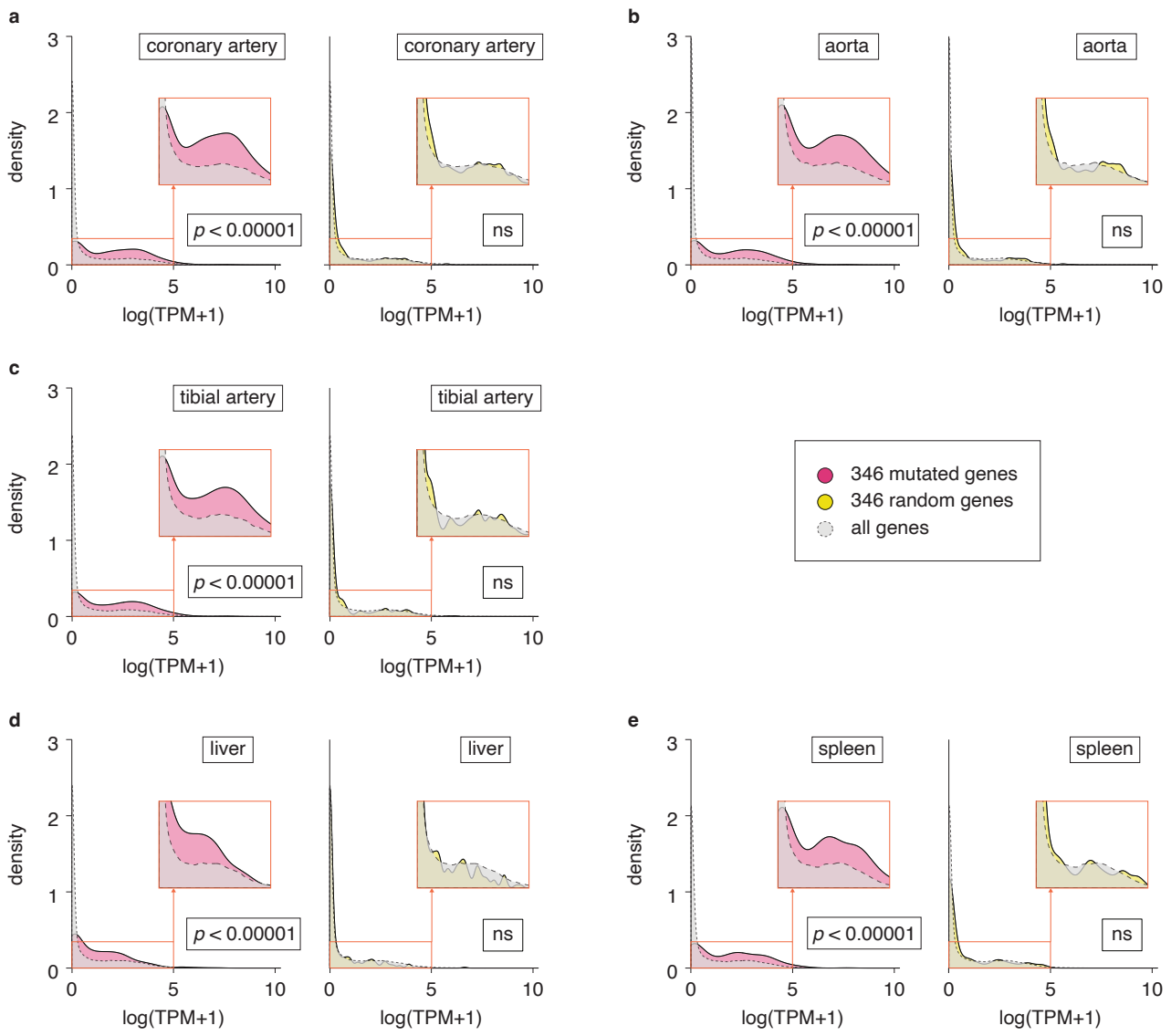
Supplementary figure S2 l

a. Variant allele frequencies (VAFs) of all plaque-confined somatic mutations (*i.e.*, mutations not identified in patient-matched buffy coats) for each patient. The horizontal dotted line shows the defined limit of detection (0.01). Dot sizes indicate variant reads, and colors indicate the type of mutation. lof = loss of function; cds = coding sequence. **b.** Density plot showing the distribution of clone cell frequencies calculated from each somatic mutation VAF from all individual patients.



Supplementary figure S3 l

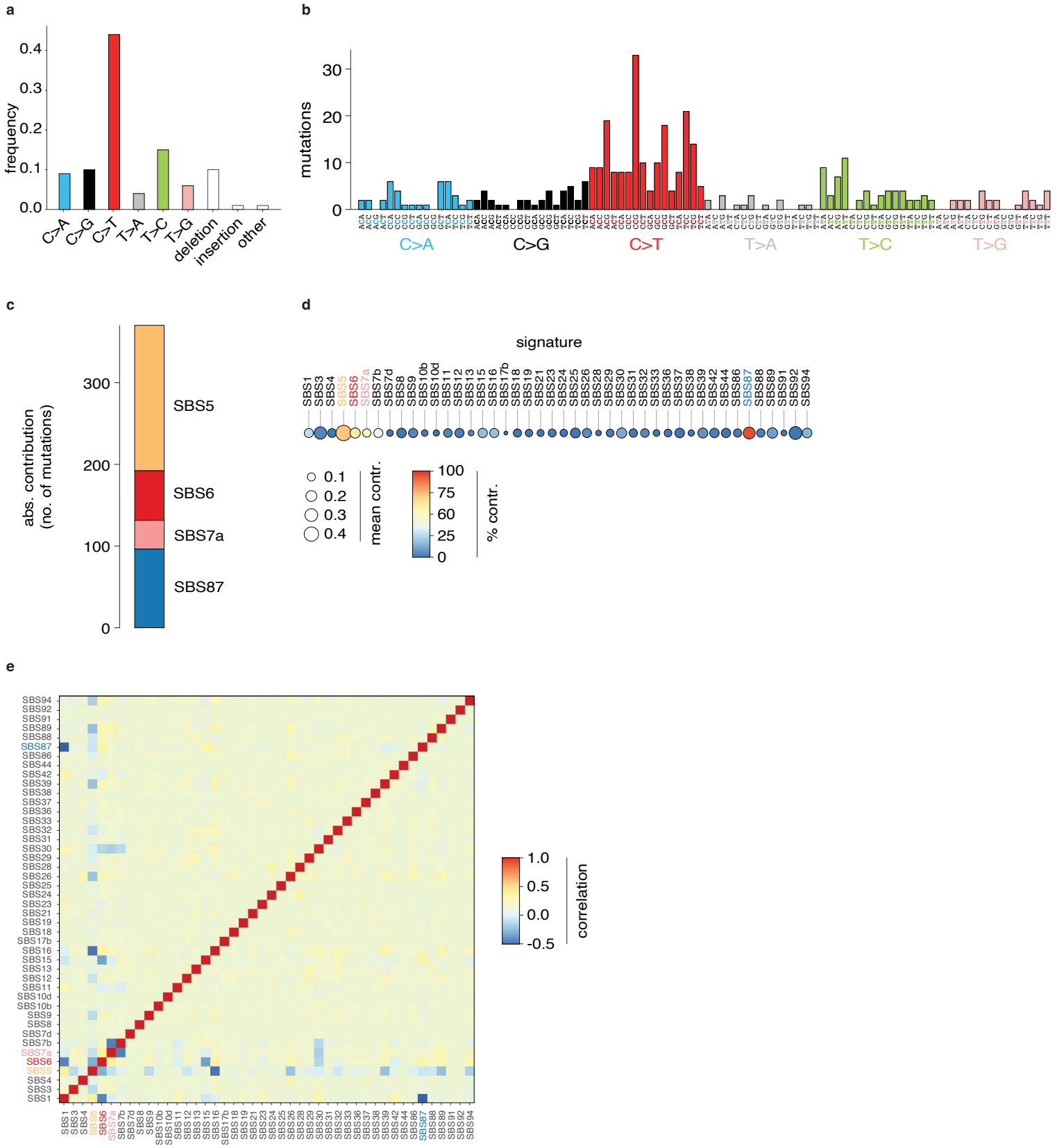
a. Upset plot showing overlaps among the top ten enriched molecular functions shown in fig. 1f. **b.** Circular network plot depicting the linkages of mutated genes with the top ten enriched molecular functions shown in fig. 1f. Sizes of dots representing molecular functions indicate the number of mutated genes attributed to each term.



Supplementary figure S4 l

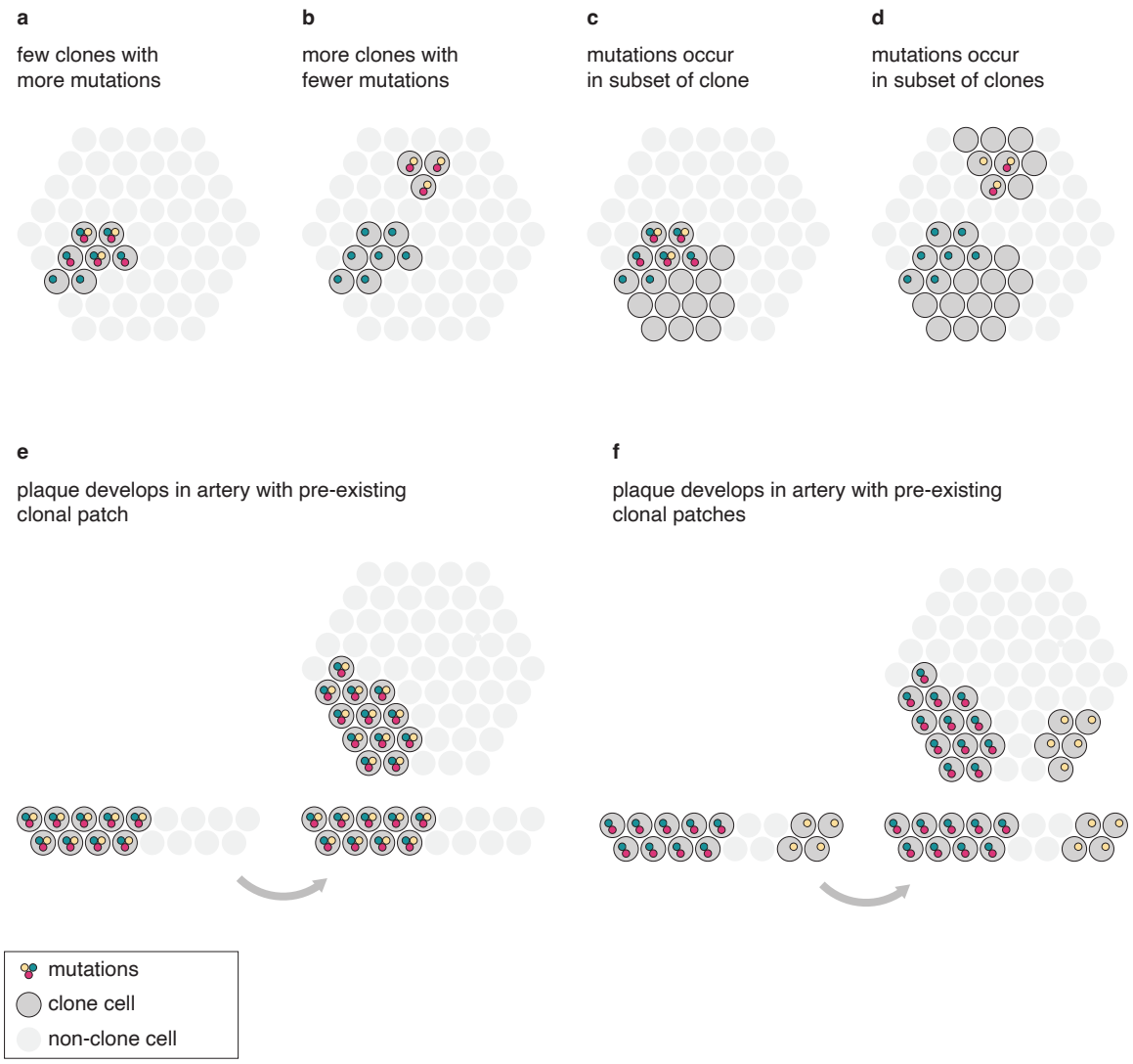
Density plot showing the distribution of gene expression levels (presented as the logarithm of transcripts per million (TPM) +1) in coronary artery- (**a**), aorta- (**b**), tibial artery- (**c**), liver- (**d**), and spleen- (**e**) tissue (data from the gene-tissue expression (GTEx) portal, v8) for 1) all genes of the genome, 2) the 346 mutated genes identified in carotid plaques, and 3) 346 genes randomly sampled from the genome. Area under the curves have been adjusted to be equal for the two datasets. To test if distributions within each graph were significantly different, a χ^2 test was performed based on the number of genes with expression levels above and below the median expression level of mutated genes or random genes, respectively.

The first graph of each panel shows that the 346 mutated genes identified in carotid plaques have significantly higher expression in all five tissue types as compared to all genes of the genome. The second graph in each panel shows that the expression level distribution of 346 randomly sampled genes would correspond to the distribution of all genes of the genome. The first panel in a is identical to fig. 1g but shown here to ease comparison with the other plots.



Supplementary figure S5 I

a. Frequency of mutations at the nucleotide level among all somatic mutations detected in carotid plaques. **b.** Catalogue of substitutions showing the frequency of trinucleotides *i.e.*, the substitution and the immediate flanking 3' and 5' nucleotides. **c.** Identified single base substitution (SBS) signatures using the strict refitting approach. **d.** Result of the bootstrapped fitting repeated 10,000 times. The identified signatures (c) all contributed to the bootstrapped solution 50% or more. **e.** Correlation between signatures. Negative correlation between signatures corresponds to their contribution being high in different bootstrap iterations. Several signatures compete for the same mutations in the refitting process due to similarities among signatures and limited number of mutations available.



Supplementary figure S6 l
Possible scenarios underlying clonal cell frequencies deduced from somatic mutations.