Medically actionable pathogenic variants in a population of 13,131 healthy elderly individuals

SUPPLEMENTARY MATERIALS

Supplementary Methods

M1. Study eligibility criteria

Each participant has lived to 70 years or older with no reported history, or current diagnosis of cardiovascular disease, cerebrovascular disease or dementia, or life-threating cancer diagnoses, as confirmed at time of enrollment into the ASPREE study (1-3). Each participant also met the following eligibility criteria at time of study enrolment; no history of a diagnosed cardiovascular disease event at baseline, defined as myocardial infarction (MI), heart failure, angina, pectoris, stroke, transient ischemic attack, 50% carotid stenosis or previous carotid endarterectomy or stenting, coronary artery angioplasty or stenting, coronary artery bypass grafting, or abdominal aortic aneurysm, no diagnosis of atrial fibrillation; no serious illness likely to cause death within the next 5 years (as assessed by general practitioner); no current or recurrent condition with a high risk of major bleeding; no anaemia (haemoglobin > 12 g/dl males, > 11 g/dl females); no current continuous use of other antiplatelet drug or anticoagulant; no systolic blood pressure ≥180 mm Hg and/or a diastolic blood pressure ≥105 mm Hg; no history of dementia or a Modified Mini-Mental State Examination (3MS) score ≤77 (4); and no severe difficulty or an inability to perform any one of the 6 Katz basic activities of daily living (ADLS)(5).

M2. Biospecimen collection

ASPREE participants contributing to the ASPREE Healthy Ageing Biobank provided opt-in consent for biospecimens to be used for biomarker and/or genetic analysis. Samples were preferentially collected prior to the start of study medication, but were allowed a collection window of up to 1 year after starting study medication. Approximately 40ml of non-fasting blood and 3ml of urine were collected from each participant and processed at 4 degrees Celsius. Blood was collected into three ethylenediaminetetraacetic acid (EDTA) containing vacutainers (10mls capacity each), two Sodium Citrate containing vacutainers (2.7ml capacity each) and one serum separator tube (SST) vacutainer (8.5ml capacity, Serum BD Vacutainer^{TM}). The EDTA vacutainer collection provided the DNA source: after centrifugation at 4 degrees Celsius, the middle white blood cell (Buffy Coat) layer was transferred into a single conical falcon tube, washed twice in red cell lysis buffer (pH7.3) and resuspended in phosphate-buffered saline (PBS). Samples were stored at -80 degrees Celsius in 250 µl aliquots until DNA extraction. Saliva samples were collected from ASPREE participants who did not contribute a blood/urine sample to the Biobank, regardless of the length of study medication exposure. A 2ml saliva sample (collected at least 30mins after eating, drinking or smoking) was collected

directly into a stabilization buffer, mixed thoroughly and stored at room temperature until DNA extraction.

M3. Library preparation, sequencing and variant calling

M3.1 Library Preparation

A custom AmpliSeq panel (Thermo Fisher Scientific) was designed inclusive of 762 genes with disease associations, Mendelian inheritance and/or used in clinical testing. The panel included ~25,000 amplicons covering 2,481,581 nucleotides of unique genomic bases in 11,593 distinct regions. Of this content, covered at the exonic level were 139 genes associated with cancer, 151 associated with cardiovascular disease, and 115 associated with pharmacogenomics. There were 116 genes with autosomal dominant (AD) inheritance, 290 with autosomal recessive (AR) and 33 with X-linked. The panel included 4259 single-amplicon 'hot-spots' for diseases-associated single nucleotide polymorphisms (SNPs) identified from genome-wide association studies (GWAS) in coding and non-coding regions.

Sample DNA was diluted to 5ng/µL and libraries were prepared with the included Ion AmpliSeq Library Kit Plus reagents and the Ion Xpress Barcode Adapters 1-16 Kit (Thermo Fisher Scientific # 4471250) to barcode each pair of samples, following standard manufacturer's instructions. Briefly, 10 ng of DNA was added to two master amplicon mixes, one for each primer pool of the DNA panel and the two primer pool reactions were amplified in different wells of the same plate following the specified cycling conditions in the user guide. Following PCR amplification, the DNA amplification reactions from the two DNA primer pool reactions were combined in wells and processed as follows. The primers were partially digested with the proprietary FuPa enzyme and each sample was barcoded with a unique IonExpress barcode. Finally, a 1.5X bead purification was performed with Agencourt AMPure XP Reagent (Beckman Coulter, Brea, CA, #A63880) following the instructions in the protocol to clean up the sample and remove adapter dimers. At the end of the cleanup protocol, the samples were eluted and amplified with a master mix of library amplification mix and Equalizer primers provided in the Ion Library Equalizer Kit (Thermo Fisher Scientific, #4482298). The amplified libraries were then incubated and washed using Equalizer kit reagents following the protocol, resulting in all the libraries normalized to ~100pMol.

M3.2 Priming the Ion Torrent sequencing chip

Following Equalization, all samples were diluted 1:2 for a final concentration of 50 pMol and 5 µl of each diluted, normalized library were combined to form a library pool and samples at a final loading concentration of 50 pMol. 25 µl of this pool were used for priming the sequencing chip. Ion Torrent chip priming and sequencing were carried out using the Ion Torrent S5XL system and the Ion Chef instrument with reagents

from the Ion 510/520/530 Kit-Chef (Thermo Fisher Scientific # A34461). Briefly, the Chef was used to bind each library DNA fragment to Ion Sphere Particles (ISPs) and clonally amplify each fragment by emulsion PCR. Amplified DNA fragments were then bound to streptavidin-coated beads and template negative ISPs were washed away. Template-bound ISPs were then prepared for sequencing by loading onto one Ion Torrent S5 530 chip for sequencing on the Ion S5XL sequencing system.

M3.3 Sequencing on the Ion Torrent Platform

The primed 530 chips were sequenced on the Ion Torrent S5TM XL System with library read length set at 400 bp and 850 flows per chip, with all other instrument settings set to the manufacturer's default for the Ion 510/520/530 Kit. Analyses of sequencing raw data were performed with Ion Torrent Suite (version 5.6.0) using the "coverageAnalysis," "sampleID," "variantCaller" (with somatic/low stringency settings for the "variantCaller"), and custom BED file plugins, with all other settings for the run report set to the manufacturer's default.

M3.4. Sequence data processing and variant calling

Torrent Suite 5.2 on Ion Torrent S5XL produced raw reads (average 200x coverage) and exported them as fastQ format. These raw reads on a per sample basis, were aligned to the human genome reference version GRCh37. Aligned bam files were passed onto the Torrent Variant Calling Suite for low-stringency variant calls. Variant calling was performed on each sample separately, which produced a single sample VCF file. Post variant calling, all per-sample VCF files were merged using bcftools merge option to produce a single multi-sample vcf file. For variant calls, which were detected in multiple samples with different depth and frequency, a default average value was used, while the original information was retained in the INFO column of the multi-sample vcf file. Each variant was annotated and the effect on the resulting protein were computing using the Variant Effect Predictor tool from Ensembl, which resulted in an annotated multi-sample vcf file.

M4. Variant curation and validation

Variants with 'pathogenic' or 'likely pathogenic' clinical annotation were identified from the ClinVar database (February 2019) (6). High-confidence predicted loss-of-function variants were identified using the Loss-Of-Function Transcript Effect Estimator (LoFTEE) tool using the 'high-confidence' filter (https://github.com/konradjk/loftee)(7). This identifies predicted Stop-gained, Splice site disrupting and Frameshift variants. Predicted deleterious missense variants without pathogenic or likely pathogenic annotations in ClinVar were excluded. All variants passing these filters were subject to technical and clinical variant curation, following best practices, including the ACMG/AMP Standards and Guidelines for the Interpretation of Sequence Variants (8).

Technical variant curation involved manual inspection of each variant in the Integrative Genomics Viewer (9). Variants were flagged for average quality scores <30 across the dataset, low reads (<50x), minor allele frequency (MAF) and/or homozygous genotype frequency >10x higher than gnomAD (Non-Finish Europeans) (7), visible evidence of mis-alignment and/or errors around the variant site, evidence of allele-specific amplification bias, imbalanced stand coverage or variant allele frequency (<25%), direct adjacency to homopolymers, and evidence of multi-allelic base calls. Variants with three or more flags were excluded as technical sequencing artefacts. Variants called as borderline were flagged for validation by Sanger sequencing. Predicted deleterious variants found to be benign, conflicting or uncertain significant (VUS) in ClinVar were excluded.

Clinical variant curation of all variants passing technical variant curation involved inspection for mechanism of disease based on the ACMG recommendations for reporting of secondary findings (10), and gene-specific curation using the following criteria. *BRCA1/BRCA2* gene variants were curated using the ENIGMA classification criteria (11), mismatch repair gene (*MLH1, MSH2, MSH6, PMS2*) variants were curated using the InSiGHT classification criteria (12), and curation of other variants was based on the ACMG/AMP guidelines (8), with modifications based on recent publications (13-16). Only pathogenic and likely pathogenic variants were included in further analyses. A total of 14 variants called as pathogenic or likely pathogenic were validated by Sanger sequencing in 26 samples, with a 100% concordance/validation rate. A further 21 variants that were called as borderline (suspected sequencing artefact) in 35 individuals were also sent for Sanger validation. Of these, 16/21 (76%) failed validation and were called negative and 5/21 (24%) were positive.

Supplementary Results

Table S1. Pathogenic variant carriers in ASPREE (per-gene)**.**

		ASPREE N=13,131
Average age – yrs (1st-3rd quartiles)		75 (72-77)
Female sex – no. (%)		7056 (54%)
Gene groups		**Number of carriers (%)**
Medically actionable genes [a]		
Autosomal dominant (N=55)		195 (1.485%)
Hereditary breast and ovarian cancer		
BRCA1		11 (0.084%)
BRCA2		22 (0.137%)
BRCA1 & BRCA2	*Female*	15/7056 (0.213%)
	Male	18/6075 (0.296%)
Lynch syndrome		
PMS2		12 (0.091%)
MSH6		7 (0.053%)
MLH1		1 (0.008%)
MSH2		0
Familial hypercholesterolemia (*pathogenic*)		
LDLR		11 (0.084%)
APOB		2 (0.015%)
PCSK9		0
Romano-Ward long QT syndrome types 1,2, 3, Brugada		
KCNQ1		22 (0.168%)
KCNH2		6 (0.046%)
SCN5A		11 (0.084%)
Hypertrophic, Dilated cardiomyopathy		
MYBPC3		15 (0.114%)
MYH7		6 (0.046%)
MYH11		0
LMNA		1 (0.008%)
MYL3		0
TNNT2		1 (0.008%)
TNNI3		1 (0.008%)
GLA		1 (0.008%)
MYL2		0
Arrhythmogenic right-ventricular cardiomyopathy		
PKP2		12 (0.091%)
DSC2		4 (0.030%)
DSG2		3 (0.023%)
DSP		3 (0.023%)
TMEM43		0
Hereditary paraganglioma-pheochromocytoma syndrome		
SDHB		3 (0.023%)
SDHC		2 (0.015%)
SDHD		0

Malignant hyperthermia *RYR1*	9 (0.069%)
Catecholaminergic polymorphic ventricular tachycardia *RYR2*	4 (0.030%)
Multiple endocrine neoplasia type 2, Familial medullary thyroid cancer, Retinoblastoma *RET*	3 (0.023%)
Familial adenomatous polyposis *APC*	1 (0.008%)
PTEN hamartoma tumor syndrome *PTEN*	1 (0.008%)
Marfan syndrome, Loeys-Dietz syndromes, and familial thoracic aortic aneurysms and dissections *FBN1*	1 (0.008%)
Li-Fraumeni Syndrome *TP53*	0
Ehlers-Danlos Syndrome Type IV *COL3A1*	0
Familial thoracic aortic aneurysms and dissections	
ACTA2	0
SMAD3	0
TGFBR1	0
TGFBR2	0
Tuberous Sclerosis Complex	
TSC1	0
TSC2	0
Multiple Endocrine Neoplasia Type I *MEN1*	0
Neurofibromatosis Type II *NF2*	0
Wilms tumor *WT1*	0
Retinoblastoma *RB1*	0
Von Hippel-Lindau Syndrome *VHL*	0
Peutz-Jeghers Syndrome *STK11*	0
Juvenile polyposis syndrome *BMPR1A*	0

[a] Genes defined by the American College of Medical Genetics (ACMG) as medically actionable for reporting of secondary findings from clinical sequencing (10).

References

1. McNeil JJ, Woods RL, Nelson MR, et al. Effect of Aspirin on Disability-free Survival in the Healthy Elderly. N Engl J Med. 2018;379(16):1499-508.
2. McNeil JJ, Wolfe R, Woods RL, et al. Effect of Aspirin on Cardiovascular Events and Bleeding in the Healthy Elderly. N Engl J Med. 2018;379(16):1509-18.
3. McNeil JJ, Nelson MR, Woods RL, et al. Effect of Aspirin on All-Cause Mortality in the Healthy Elderly. N Engl J Med. 2018;379(16):1519-28.
4. Teng EL, Chui HC. The Modified Mini-Mental State (3MS) examination. The Journal of clinical psychiatry. 1987;48(8):314-8.
5. Katz S, Akpom CA. A measure of primary sociobiological functions. International journal of health services : planning, administration, evaluation. 1976;6(3):493-508.
6. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic acids research. 2014;42(Database issue):D980-5.
7. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.
8. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in medicine : official journal of the American College of Medical Genetics. 2015;17(5):405-24.
9. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.
10. Kalia SS, Adelman K, Bale SJ, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genetics in medicine : official journal of the American College of Medical Genetics. 2017;19(2):249-55.
11. Spurdle AB, Healey S, Devereau A, et al. ENIGMA--evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. Hum Mutat. 2012;33(1):2-7.
12. Thompson BA, Spurdle AB, Plazzer JP, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat Genet. 2014;46(2):107-15.
13. Kelly MA, Caleshu C, Morales A, et al. Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. Genetics in medicine : official journal of the American College of Medical Genetics. 2018;20(3):351-9.
14. Abou Tayoun AN, Pesaran T, DiStefano MT, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. Hum Mutat. 2018;39(11):1517-24.
15. Oza AM, DiStefano MT, Hemphill SE, et al. Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. Hum Mutat. 2018;39(11):1593-613.
16. Mester JL, Ghosh R, Pesaran T, et al. Gene-specific criteria for PTEN variant curation: Recommendations from the ClinGen PTEN Expert Panel. Hum Mutat. 2018;39(11):1581-92.