Penetrance and the Healthy Elderly

Paul Lacaze,¹ Ingrid Winship,^{2,3} and John McNeil¹

The variable penetrance of pathogenic variants (PVs) represents a major challenge to the field of human genetics, often complicating clinical decision-making and risk management. Nonpenetrance, the detection of PVs in the absence of disease manifestation, is a common phenomenon, yet, we know very little about the underlying factors, which may protect some individuals and not others. Placing a new focus on the genomic study of the healthy elderly may be pivotal for advancing our understanding of penetrance. Studying those who remain unaffected late into life, despite harboring known genetic risk variants, could provide important insights into disease mechanisms and ultimately inform clinical care, yet, it has received relatively little attention as a research strategy. The ever increasing use of sequencing technology is further driving the requirement to understand the penetrance of ascertained variants. The ASPREE Biobank of Healthy Ageing provides a unique opportunity to address this area of need. DNA has been collected from a cohort of over 14,000 healthy elderly individuals aged 70 years or older enrolled in an aspirin clinical trial. The ASPREE cohort represents a healthy reference population ascertained without the typical biases of a genetic study. The cohort is depleted of expressed monogenetic disease, yet will contain hundreds of elderly individuals with known PVs in clinically actionable genes. Investigating this population along with other cohorts of the healthy elderly will provide critical new knowledge into the penetrance of actionable variants as a foundation for informing clinical care.

Keywords: healthy elderly, penetrance, genetics

REDUCED PENETRANCE OCCURS when only a portion of the total individuals carrying a genetic variant in the population develop the associated trait or disease. Many single-gene disorders and predispositions show reduced penetrance, causing challenges for the clinical utility of pathogenic mutations in these genes. Even for genetic variants with well-established clinical actionability, penetrance varies significantly across the population, complicating decisions for intervention based on genotype alone, without family history or disease symptoms. Penetrance can vary substantially among ethnicities, by age, and in response to environmental factors (Cooper *et al.*, 2013). Our inability to explain nonpenetrance reflects a gap in our understanding of causative genetic variation, ultimately constraining the use of genetic data for clinical decision-making.

Why reduced penetrance occurs, or conversely, the underlying mechanisms of protection that may exist for those individuals unaffected, remains unclear. Rarely has focus been placed on identifying elderly nonpenetrant individuals

who lack typical disease symptoms late into life, despite carrying a pathogenic variant (PV). However, these individuals are likely to be extremely valuable for advancing our understanding of penetrance.

To date, there has been an absence of genuine representative population data to calculate penetrance, a major issue for the field. Part of the reason for this is the lack of well-phenotyped reference populations of confirmed *unaffected* individuals known to be harboring pathogenic mutations late into life in the absence of typical disease features. Penetrance estimates therefore have been suboptimal due to limited sample sizes and a focus on affected families undergoing genetic testing, rather than on healthy individuals or more randomly ascertained sample populations.

Surveys of ostensibly healthy adults, more representative of the population average, have only recently begun at scale (Dewey *et al.*, 2016; Lek *et al.*, 2016; Telenti *et al.*, 2016). This has been due to reductions in cost of DNA sequencing and the availability of biobank sample cohorts not

¹Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia.

²Genetic Medicine and Family Cancer Clinic, Royal Melbourne Hospital, Melbourne, Australia.

³Department of Medicine, University of Melbourne, Royal Melbourne Hospital, Melbourne, Australia.

ascertained using traditional genetic strategies. Observations from some of the first sequencing analyses of “reference” human populations are beginning to improve our understanding of penetrance and variant pathogenicity, yet, far more progress is required. The Exome Aggregation Consortium (ExAC) has perhaps been the most notable, taking the important step toward generating *objective* metrics of pathogenicity informed by population allele frequency (Lek *et al.*, 2016).

Other efforts have identified highly penetrant PVs in clinically actionable genes in otherwise healthy or unaffected adults (Dewey *et al.*, 2016; Lek *et al.*, 2016; Telenti *et al.*, 2016), sometimes in the absence of any disease features (Natarajan *et al.*, 2016). Additional evidence even suggests rare cases of “resilience” against pathogenic mutations previously thought to be fully penetrant or causative of severe childhood-onset diseases (Chen *et al.*, 2016), challenging the very notion of genetic determinism.

These recent efforts have been pivotal for achieving early gains from genetic surveys of human reference populations toward understanding penetrance. However, many of the conclusions made from these studies have been limited by the opportunistic nature of data collection and/or sample aggregation (Chen *et al.*, 2016; Dewey *et al.*, 2016; Lek *et al.*, 2016; Natarajan *et al.*, 2016; Telenti *et al.*, 2016), often characterized by one or more of the following limiting factors;

- (1) Merging of datasets using control samples from different case/control studies, ascertained in different contexts, rather than starting with a well-defined reference set of healthy elderly individuals confirmed to be free of disease late into life.
- (2) Variable types of genomic data collection, sometimes involving different genotyping or sequencing platforms and/or variant calling methods, which can limit data consistency and quality (Chen *et al.*, 2016).
- (3) A lack of detailed or systematic phenotyping across reference populations to confirm or negate the presence or absence of disease features with confidence.
- (4) Variable age of collection for adult samples, ranging from older individuals to younger individuals, possibly younger than the age of onset for some genetic conditions.
- (5) Absence of validation, rephenotyping and clinical follow-up of genetic findings to interpret penetrance due to restricted consent models prohibiting re-contact, a common limitation of genetic analyses.

Importance of the Healthy Elderly

Studies of the healthy elderly could be instrumental for advancing our understanding of penetrance. In particular, identifying those elderly individuals who carry PVs, but remain disease-free late into life, will provide the unique opportunity to begin to create a framework for studying mechanisms of protection against pathogenic variation. Such individuals could be more confidently described as nonpenetrant or “resilient” due to their advanced age, but will nonetheless require deep clinical phenotyping to assign the presence or absence of disease features with certainty. Identification and clinical assessment of these individuals provide a novel and potentially informative approach to the

mechanistic exploration of nonpenetrance; this would include deep phenotyping, family studies, disease-specific clinical assessments, whole-genome sequencing of nonpenetrant individuals, and even functional studies of patient-derived cell lines—a new paradigm of research focus.

The ASPREE Biobank of Healthy Ageing provides a unique opportunity to undertake this, with biospecimens collected from over 14,000 healthy elderly individuals aged 70 years or older enrolled in an international aspirin clinical trial (Group, 2013; McNeil *et al.*, 2017). The biospecimens have been collected from ASPREE study participants for research purposes, to enable various kinds of biomarker studies and genetic analyses to accompany the primary trial.

The study entry criteria for participation in the ASPREE trial included no signs of life-threatening cardiovascular disease, cancer, or cognitive decline at time of enrollment [8], hence, ASPREE can be considered a healthy aging cohort. ASPREE participants also had to meet further entry criteria, including the absence of any serious illness likely to cause death within the next 5 years, conditions with high risks of major bleeding, diagnosed cardiovascular events, including atrial fibrillation, history of anemia, current use of antiplatelet drug or anticoagulants, high blood pressure, or severe physical disability (Group, 2013).

Therefore, ascertainment of the ASPREE sample population was determined by strict trial inclusion criteria, not by the traditional ascertainment methods of a genetic study (e.g., affected families or case–control studies). This ascertainment method, combined with the advanced age of the cohort, makes ASPREE highly atypical and provides unique circumstances for the study of gene penetrance.

In addition, each ASPREE participant has been phenotyped systematically over a period of 3–7 years of study participation—capturing medical record data, health history, family history of disease, hospitalizations, and other relevant clinical events during the ASPREE trial (Group, 2013; Lacaze *et al.*, 2017). All primary and secondary endpoints of the study, which include cardiac events, strokes, cancers, dementia, disability, and clinically significant bleeding have been adjudicated by expert panels. The ASPREE population therefore enables detection of a PV to be interpreted within a rich context of phenotypic, medical history and clinical trial data across thousands of consented biobank samples, all from confirmed healthy elderly, nonsymptomatic individuals. This will be highly effective for identifying and interpreting apparent cases of life-long nonpenetrance to known genetic risk factors used routinely in predictive genetic testing (Table 1 for cohort feature summary).

Rather than take the traditional disease or phenotype-driven approach to genomic analyses, the ASPREE Biobank can take the converse approach of screening thousands of consented samples for the presence of known PVs in clinically actionable genes, then mine-rich available phenotypic data for evidence of related disease features or family history to interpret penetrance. We estimate the rate of PVs in clinically actionable genes (based on the American College of Medical Genetic list (Kalia *et al.*, 2017) will be around 1.5–3.5% according to previously published observations in smaller reference populations (Dorschner *et al.*, 2013; Song *et al.*, 2016)). In ASPREE, this could amount to ~200–500 healthy elderly individuals harboring actionable PVs from a total sample size ~14,000 samples.

TABLE 1. UNIQUE FEATURES OF THE ASPREE BIOBANK OF HEALTHY AGING

Feature	Details
Healthy, elderly reference population, ascertained without traditional biases	Over 14,000 DNA samples collected from healthy elderly research participants enrolled in an aspirin clinical trial, all above the age of 70 years with no major symptoms of life-threatening genetic disease late into life. Ascertained based on age and trial inclusion criteria only, not genetic study design.
Potential for confirmed case of life-long nonpenetrance	ASPREE participants with known actionable pathogenic variants can be identified at an advanced age with the absence of related disease features. These can provide high-confidence cases of life-long nonpenetrance or “resilience” to known pathogenic mutations
Recontact consent, potential for follow-up studies	ASPREE biobank participants consented to genetic analyses under the agreement findings of personal or familial health significance may be returned to them by a health professional if discovered during research. This allows scope for recontact and follow-up studies, including deeper phenotyping and potential recruitment of family members
Systematic, longitudinal phenotyping	Enrollment in a clinical trial means medical records, hospitalizations and clinical endpoints of the ASPREE trial, have been collected systematically over time to a high standard. Additional measures such as cognitive assessments, physical function tests, and family histories have also been collected systematically and longitudinally
Limited ethnic diversity	Predominately a Caucasian cohort of limited mixed ancestry (disadvantageous for overall understanding of human genetics, but advantageous for interpretation of pathogenic variants derived predominately from Caucasian populations)

We expect variants to be distributed roughly according to disease prevalence of actionable genetic conditions, with variants related to hereditary breast and colorectal cancer and cardiomyopathy to be the most common. However, it is possible that some, or even all, allele frequencies for monogenic disease may be depleted in ASPREE due to the selection bias toward healthy elderly. Alternatively, allele frequencies may be equivalent to younger populations, but within a selection bias toward nonpenetrant individuals.

Consent for the ASPREE Biobank also encompasses possible recontact and scope for follow-up studies, as biobank participants consented to receive medically significant genetic research findings if identified (Lacaze, 2017). Actionable PVs, even if found in nonpenetrant elderly individuals, could be of medical significance to younger family members carrying the same often autosomal dominant mutation who may be subsequently affected. A new research study is therefore being designed to examine genetic findings from the ASPREE Healthy Ageing Biobank, which will provide actionability criteria for the elderly, inform decisions of disclosure, capture medical events postdisclosure, and allow potential family recruitment as well as studies focused on penetrance.

We propose five possible methods/approaches for analyzing penetrance in the cohort;

- (1) Calculating “aggregate penetrance” for groups of genes together, rather than on a per-gene basis due to small sample numbers for any one gene (Natarajan *et al.*, 2016). This would inform the combined penetrance of PVs within groups of related genes across a population, for example, cancer syndrome genes versus cardiac disease genes.

- (2) Family history data to analyze the penetrance of autosomal dominant PVs across families. Detailed family history data are available (mother, father, sibling, and child) collected at baseline and divided into cancer subtypes and cardiovascular disease subtypes, as well as stroke, dementia, and other comorbidities. This can be mined for the presence/absence of related disease family history, categorized per-gene, per-disease, or per-disease category to inform penetrance. This will determine, for example, whether certain genes in the ACMG list manifest in higher rates of family history of disease than others, in this cohort.
- (3) Individual patient medical histories (self-reported at baseline) provide many fields which contain potential evidence of related clinical features/phenotypes to confirm the presence or absence of penetrance. This includes personal cancer histories broken down by cancer subtype, cardiovascular histories, including blood pressure, hypertension, cholesterol, heart attacks, and other relevant fields such as BMI which can be mined. No detectable evidence of any related phenotype in these fields provides evidence of nonpenetrance, also a significant finding, measured in a systematic way.
- (4) Longitudinal clinical outcome data (when available, still being collected). Penetrance can be recalculated if PVs become penetrant over the course of the ASPREE study period (3–7 years). This would only be relevant for the latest onset diseases, which appear beyond the age of 70, or for subclinical-related disease features which may manifest in that time.
- (5) Disclosure of findings and family studies. The ultimate method to study penetrance would be to disclose

the genetic findings (PVs), then, work toward re-phenotyping individuals and their family members to construct pedigrees. This would require significant time and effort with informed consent, but would be most powerful.

A new focus toward the genetics of the healthy elderly will help identify individuals who are inherently protected against the pathogenic effects of known genetic risk factors. However, to discover and understand the underlying mechanisms explaining nonpenetrance will be more challenging and require additional studies. It may include phenomena such as epigenetic allele silencing in the tissues associated with the expression of disease phenotype, or genetic mosaicism, whereby copies of either the maternal or paternal allele contribute to disease risk or tissue-specific expression. Buffering variants found throughout the genome are also likely to contribute to risk and influence penetrance. The concept of generating polygenic scores of “resilience” (i.e., the many hundreds or thousands of buffering factors that may contribute to expression of a known risk factor) may also eventually be required to deliver more precise predictive testing and risk management.

A far more sophisticated understanding of penetrance is essential for realizing the full potential of genomic medicine. The healthy elderly who have apparent protection against the pathogenic effects of known mutations may provide an important window into the current mystery of penetrance. We have made progress toward the concept of *measuring* penetrance, rather than just *estimating* it. The next step will be studying it, *understanding* it through discovery and annotation of buffering mutations and protective factors, and then, addressing it clinically with more data-driven and evidence-based approaches.

Author Disclosure Statement

No competing financial interests exist.

References

- Chen R, Shi L, Hakenberg J, *et al.* (2016) Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nat Biotechnol* 34:531–538.
- Cooper DN, Krawczak M, Polychronakos C, *et al.* (2013) Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet* 132:1077–1130.
- Dewey FE, Murray MF, Overton JD, *et al.* (2016) Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* 354:pii: aaf6814.
- Dorschner MO, Amendola LM, Turner EH, *et al.* (2013) Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am J Hum Genet* 93:631–640.
- Group AI (2013) Study design of ASPIrin in Reducing Events in the Elderly (ASPREE): a randomized, controlled trial. *Contemp Clin Trials* 36:555–564.
- Kalia SS, Adelman K, Bale SJ, *et al.* (2017) 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. *Genet Med* 19:249–255.
- Lacaze P, Winship I, Ryan J, McNeil JJ (2017) Pathogenic variants in the healthy elderly: unique ethical and practical challenges. *J Med Ethics* [Epub ahead of print]; DOI: 10.1136/medethics-2016-103967.
- Lacaze P, Woods R, Zoungas S, *et al.* (2017) The genomic potential of the Aspirin in Reducing Events in the Elderly and Statins in Reducing Events in the Elderly studies. *Intern Med J* 47:461–463.
- Lek M, Karczewski KJ, Minikel EV, *et al.* (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536:285–291.
- McNeil JJ, Woods RL, Nelson MR, *et al.* (2017). Baseline Characteristics of Participants in the ASPREE (ASPIrin in Reducing Events in the Elderly) Study. *J Gerontol A Biol Sci Med Sci* [Epub ahead of print]; DOI: 10.1093/gerona/glw342.
- Natarajan P, Gold NB, Bick AG, *et al.* (2016). Aggregate penetrance of genomic variants for actionable disorders in European and African Americans. *Sci Transl Med* 8:364ra151.
- Song W, Gardner SA, Hovhannisyan H, *et al.* (2016) Exploring the landscape of pathogenic genetic variation in the ExAC population database: insights of relevance to variant classification. *Genet Med* 18:850–854.
- Telenti A, Pierce LC, Biggs WH, *et al.* (2016) Deep sequencing of 10,000 human genomes. *Proc Natl Acad Sci U S A* 113: 11901–11906.

Address correspondence to:
Paul Lacaze, PhD

Public Health Genomics Program
Department of Epidemiology and Preventive Medicine
School of Public Health and Preventive Medicine
Monash University, The Alfred Centre
99 Commercial Road
Melbourne VIC 3004
Australia

E-mail: paul.lacaze@monash.edu