Decoding telomere gene variation through saturation genome editing: from functional genomics to clinical interpretation

Obolenski, S.

8

General Discussion

SUMMARY OF MAIN FINDINGS AND THESIS CONTRIBUTIONS

Interpreting rare germline variants in cancer susceptibility genes remains a major challenge, particularly for genes involved in telomere maintenance, given their capacity for context-dependent, bimodal effects on telomere length regulation. As hereditary cancer testing panels expand, increasing numbers of variants are identified in genes such as *POT1*, *ACD* and *TINF2*, however for many of these there is limited functional or clinical evidence[1,2]. This results in a high burden of variants of uncertain significance (VUS), which complicate molecular diagnosis, clinical risk stratification and personalized surveillance recommendations[3–5].

This thesis addresses this problem by evaluating saturation genome editing (SGE) as a scalable experimental strategy for characterizing variants in multiplex in genes linked to cancer predisposition with long telomeres (CPLT) disorders[6]. Using *POT1* and *TINF2* as model genes, the work presented here combines high-throughput variant maps with genetic, clinical and population data to improve genetic risk interpretation in hereditary cancer.

The first major contribution of this thesis is the implementation and application of SGE on telomere biology genes in a human haploid cell model. **Chapter 2** positions SGE within the broader framework of multiplexed assays of variant effect (MAVE) while **Chapter 3** describes the optimization of an SGE workflow in HAP1-A5 cells. This includes variant library and sgRNA design, genome-editing strategies and the use of quantitative fitness-based readouts for genes in which loss-of-function (LoF) affects HAP1 cellular viability[7,8]. Building on this platform, **Chapters 4** and **5** apply SGE to generate high-resolution functional maps for *POT1* and *TINF2*. For *POT1*, 13,919 variants were assayed at the endogenous locus, while for *TINF2*, 4,261 variants in exons encoding key shelterin-interaction domains were assayed. Together, these maps define regions of functional constraint and tolerance to variation across both genes.

The second contribution of this thesis is the correlation of SGE derived data with telomere phenotypes and clinically annotated cancer diagnoses, as described in **Chapters 4** and **5**. For *POT1*, SGE scores were associated with telomere length and cancer incidence in UK Biobank (UKBB), supported by additional analysis of patient-derived samples as well as experimental telomere length and telomeric DNA damage repair (DDR) signalling validation in the HAP1-A5 cell model. For *TINF2*, functional SGE scores were evaluated using the available clinical evidence to date, together with novel variants from cancer-prone families. Together, these

analyses illustrate how high-throughput functional data can be integrated with genetic and clinical evidence to support variant interpretation.

The third contribution of this thesis is the evaluation of SGE data in both population-based and clinically ascertained melanoma cohorts. **Chapter 6** shows that pathogenic *POT1* variants account for a small proportion of melanoma at the population level but remain clinically relevant in selected high-risk individuals. Functional assays, including SGE, identify only a subset of rare variants as functionally disruptive, supporting targeted testing rather than population-wide screening. **Chapter 7** extends this analysis to Swedish melanoma families and confirms that pathogenic *POT1* variants are rare but clinically informative. On this basis, *POT1* testing is recommended for families with multiple melanomas, early-onset disease, spitzoid histopathology, or co-occurring syndromic malignancies.

Taken together, this thesis demonstrates how high-throughput functional SGE data can be interpreted in the context of hereditary cancer genetics. It shows that SGE can be applied to telomere maintenance genes, that the resulting scores capture biologically meaningful differences between variants, and that these data can be used to support variant interpretation when integrated with genetic, clinical and population-level evidence.

FUNCTIONAL INSIGHTS FROM SATURATION GENOME EDITING OF TELOMERE GENES

A central outcome of this thesis is the demonstration that SGE, as a form of MAVE, can be applied effectively to clinically relevant telomere biology genes. The functional maps generated for *POT1* and *TINF2* delineate regions of the shelterin complex that are particularly sensitive to genetic variation, highlighting the importance of intact telomere regulation for normal cellular fitness.

POT1 **-** A SPECTRUM OF FUNCTIONAL ALLELES AND ASSOCIATED CANCER RISK

In **Chapter 4** of this thesis a comprehensive functional analysis of coding variation in *POT1*, a key component of the shelterin complex with established roles in telomere protection[9] and inherited cancer risk[10], is presented. By applying SGE at the endogenous genomic locus, 97% of all possible sequence single nucleotide variants (SNVs) were systematically interrogated in a uniform experimental context. The resulting dataset enables evaluation of *POT1* function at nucleotide resolution and provides a framework for variant interpretation.

One of the central findings of **Chapter 4** is that *POT1* variants do not segregate into a simple pathogenic-benign dichotomy. Instead, variants span a continuum of functional effects, allowing classification into strong and weak depleted, as well as functionally neutral / unchanged and enriched categories. This functional stratification shows that missense variants exert graded impacts on *POT1* function, which is a critical observation since many clinically observed *POT1* variants are missense. Importantly, SGE functional scores correlated with estimated cancer risk in the UKBB analysis, with strong depleted *POT1* variants associated with a disease odds ratio of approximately 5.6 and weak depleted variants showing a more modest increase in risk (odds ratio ~2.5). These findings suggest that cancer risk associated with *POT1* variation may reflect the degree of functional impairment rather than a strict binary threshold, with partially functional alleles conferring intermediate risk. This pattern is consistent with previous MAVE studies[11,12], in which SGE of *DDX3X* and *RAD51C* similarly distinguished fast-/strong depleted variants, enriched for truncating and core-domain missense alleles, from more slow depleted missense variants. In those studies, the slowly depleted variants were interpreted as hypomorphic alleles retaining partial function[12]. As can be seen across *POT1*, *DDX3X* and *RAD51C*, SGE in the HAP1-A5 cell model can robustly delineate multiple levels of functional impairment, with classifications that align closely with available genetic and clinical data. For *POT1* in particular, SGE fitness effects aligned with telomere-associated phenotypes observed in human datasets, including UKBB and patient-derived samples. Depleted variants were associated with telomere elongation in both human carriers and edited HAP1-A5 cells, accompanied by increased telomeric *53BP1* foci, consistent with impaired telomere end protection and ataxia telangiectasia and Rad3-related (ATR)-mediated DDR. Together, these observations support the conclusion that SGE fitness effects provide a biologically meaningful readout of *POT1* dysfunction across cellular and human contexts[13,14].

Mechanistically, the association of depleted variants with telomere elongation is consistent with loss of POT1-mediated restraint of telomerase access at the single-stranded telomeric overhang[9,15,16]. Within the shelterin complex, POT1–TPP1 both protects the telomere end and regulates telomerase recruitment and activity; disruption of this interaction can permit increased telomerase access and progressive telomere extension despite impaired end protection[17]. This interpretation is consistent with models in which POT1 normally competes with replication protein A (RPA) for binding to single-stranded telomeric DNA and, through its tethering to TPP1 within shelterin, limits access of telomerase to the chromosome terminus[15]. In contrast, variants that more severely compromise POT1-dependent end protection or replication-associated functions, such as regulation of C-strand fill-in, would be expected to

promote telomere instability and shortening, although such phenotypes are not the dominant signal observed in this SGE system.

Mapping functional SGE data onto POT1 structure further supported the biological relevance of these findings. Strong and weak depleted variants clustered within the DNA-binding oligonucleotide/oligosaccharide-binding (OB) domains and at highly conserved residues, whereas enriched variants tended to localise to less conserved positions outside key interaction surfaces. In addition, the depletion of nearly all upstream truncating variants, with few unchanged truncations restricted to the terminal exon, is consistent with nonsense-mediated mRNA decay (NMD) of upstream alleles and tolerance of C-terminal loss, as reported in other SGE studies[11,12,18,19]. Together, these observations confirm that SGE depletion reflects domain-specific loss of POT1 function rather than generic fitness effects.

Importantly, the *POT1* SGE map identified a subset of synonymous variants that were significantly depleted and correlated with high SpliceAI scores, indicating likely aberrant mRNA splicing effects[20]. Consistent with previous SGE studies[12,18,19], these results demonstrate that nucleotide-resolution SGE can capture RNA-mediated functional consequences, including altered splicing and transcript abundance, in addition to protein-level effects[18]. The clinical relevance of these findings is exemplified by the identification of a melanoma patient carrying a synonymous but strong depleted *POT1* variant (Q249=) with a notable family history of melanoma. Collectively, these observations underscore the importance of considering non-obvious functional consequences when interpreting ostensibly ‘silent’ variants in telomere-associated genes.

Overall, the *POT1* SGE data demonstrate that functional variation exists along a continuum linking graded disruption of POT1 to telomere dysregulation and cancer risk across human cohorts. Together, these findings support the use of SGE as an experimentally grounded, complementary approach for interpreting rare germline variants in telomere-associated cancer genes.

TINF2: SGE ON CRITICAL PROTEIN DOMAINS AND FUNCTIONAL EFFECTS

In **Chapter 5**, the SGE approach was extended to *TINF2*, demonstrated that variants affecting the core shelterin-binding domains are consistently associated with significant functional depletion, aligning with known pathogenic alleles[21–25] as well as newly identified cancer-associated variants. These findings indicate that SGE provides experimentally grounded support for pathogenicity assessment in *TINF2*, particularly where clinical or *in silico* evidence alone is inconclusive. Clinical correlation analyses also indicated that germline *TINF2* variants

are associated with a broad tumour spectrum, with melanoma and non-medullary thyroid carcinoma (NMTC) emerging as core malignancies, supporting routine surveillance even in the absence of a family history, in line with previous clinical recommendations by Jensen *et al*[22]. A major additional finding was the recurrent occurrence of rare malignancies, including sarcoma, renal cell carcinoma (RCC), and brain tumours, across independent *TINF2* variant carriers, which, together with prior reports, supports their recognition as clinically relevant components of the *TINF2*-associated cancer spectrum and consideration of targeted surveillance in affected families[21–24,26]. Beyond defining the disease spectrum, the *TINF2* SGE dataset also supports re-evaluation of current clinical variant classifications, as several variants annotated as VUS or with conflicting interpretations showed clear functional depletion, indicating that selected cases may warrant reassessment. Given the rarity and phenotypic heterogeneity of pathogenic *TINF2* variants, continued accumulation of well-annotated clinical data will be essential to refine interpretation, and a complete SGE map covering all coding exons and isoforms would be particularly valuable.

Structural mapping showed that depleted clinical variants cluster at the TRF2-, TPP1 and TRF1-binding interfaces, supporting the interpretation that disease-associated effects arise from disruption of key shelterin protein-protein interactions and are detectable as fitness defects in SGE[6,14]. Consistent with this, *TINF2* appeared more permissive to missense variation than *POT1*, with deleterious and significantly depleted variants concentrated at functionally critical interaction sites, while many other missense substitutions were functionally neutral, indicating position-specific tolerance within these domains. One explanation for this difference may lie in the distinct roles of these proteins within the shelterin complex: POT1 uniquely binds single-stranded telomeric DNA[13,14,27], making variants in its OB fold domains more likely to disrupt DNA binding and telomere capping, whereas TIN2 functions primarily as a scaffold stabilising shelterin assembly and may therefore tolerate a broader range of substitutions. These functional differences likely contribute to the observed variation in mutational sensitivity between the two proteins.

Overall, the *TINF2* screen demonstrates that SGE can identify both clearly disruptive alleles and clinically reported missense variants with LoF effects. It also reveals domain-specific differences in mutational tolerance and highlights the importance of integrating functional, structural and clinical information when assessing variant pathogenicity. Together, these findings support the broader applicability of SGE as functional evidence for variant interpretation in telomere biology genes.

INTEGRATING SGE INTO CLINICAL VARIANT INTERPRETATION FRAMEWORKS

To be clinically meaningful, SGE data must be interpreted within established variant classification frameworks rather than considered in isolation[28,29]. In clinical practice, germline VUS are usually classified using the American College of Medical Genetics and Genomics (ACMG) / Association for Molecular Pathology (AMP) guidelines, which integrate multiple types of evidence (e.g. population data, segregation, computational prediction, and functional studies) and weight them by evidence strength (supporting, moderate, strong or very strong)[30]. Under the original formulation of this framework, evidence from MAVE data is applicable only at a strength of strong (PS3/BS3), provided assays are "well-established"; that is, "validated and shown to be reproducible in a clinical diagnostic laboratory setting"[28,30]. However, the framework did not provide guidance on assessing the quality of an assay's validation and reproducibility, rendering the application of PS3/BS3 highly inconsistent between diagnostic centres[28,29].

To expand upon the original 2015 recommendations for application of PS3/BS3, the Clinical Genome (ClinGen) Resource, through the work of the Sequence Variant Interpretation (SVI) Working Group, developed general guidance for evaluating functional assays for clinical use[31]. While this guidance established a principal framework for assessing assay validity and evidence strength, it was built around traditional, low-throughput assays and did not fully address the implementation challenges of MAVE, including large-scale validation, clinical truthset construction, and quantitative evidence calibration. The Brotman Baty Mutational Scanning Working Group have subsequently addressed these challenges in their MAVE-specific recommendations[28]. Complementing this, Brnich *et al.*[32] also introduced a quantitative likelihood-based framework for calibrating functional assays against clinical truthsets and mapping results to ACMG/AMP evidence strengths, and providing recommended minimum truthset size.

In practice, the limited availability of high-confidence clinical truthsets means that meeting the requirements of the Brnich framework is challenging for many MAVE datasets[28,30]. These datasets are therefore unable to attain the PS3 (pathogenic, well-established damaging effect) / BS3 (benign, well-established lack of damaging effect) evidence criteria. Further, variability in assay design and performance and absence of standardised calibration methods means that inconsistency ultimately remains in the precise application of PS3/BS3 criteria across laboratories[28,29].

Using the Brnich framework, ClinGen Variant Curation Expert Panels (VCEPs) have developed gene-specific ACMG variant classification guidelines and, for genes such as *BRCA1*[33], *TP53*[34] and *PTEN*[35], have formally reviewed specific MAVE assays and, where appropriate, recommended their use as functional evidence in clinical variant interpretation. Notably, these same genes have recently also been used to show that MAVEs provide clinical value beyond VUS reclassification.

MAPPING SGE SCORES TO ACMG/AMP AND CLINGEN FRAMEWORKS

Chapter 4 of this thesis tested whether *POT1* SGE data could be used for ACMG/AMP variant classification. Using curated clinical truthsets and the quantitative framework developed by Brnich *et al.*[32], SGE scores were translated into ACMG/AMP PS3 and BS3 evidence strengths. The assay achieved a likelihood ratio of ~15 in favour of pathogenicity (near-strong evidence) and >190 in support of benign classification (strong evidence), indicating that strong depleted variants support pathogenic classification while unchanged SGE scores support benign interpretation. Weak depleted were excluded from calibration because they likely represent intermediate effects and are not compatible with the fully penetrant disease model under which truthsets were constructed. Overall, these findings demonstrate that *POT1* SGE satisfies key VCEP criteria for calibrated functional evidence and can be interpreted within the established ACMG/AMP framework[28,30,32]. Similar results have been reported for other cancer predisposition genes, such as *BAP1*[19] and *RAD51C*[12], where SGE data have also been formally evaluated under the ACMG framework. Of note, the *TINF2* SGE dataset could not be calibrated via the ACMG/AMP guidelines, due to very limited availability of high-confidence pathogenic and benign variants, reflecting the challenges associated with small genes and rare variants. This illustrates an important practical constraint of applying the Brnich framework *et al.*[32] to MAVE datasets: its utility depends not only on assay performance, but also on the availability of a sufficiently sized and well-curated clinical truthset spanning both pathogenic and benign variation. In the case of *POT1*, the combination of stronger clinical annotation, a larger number of observed coding variants, and the availability of variants that could be classified with confidence made quantitative calibration feasible. By contrast, for *TINF2*, the small gene size, rarity of clearly classified variants, and limited benign comparator set restricted formal calibration, even though the functional dataset itself was informative. Thus, the limiting factor for ACMG/AMP-style analysis is often not the generation of high-quality MAVE data per se, but the minimum volume and quality of independent clinical evidence required to anchor assay scores to pathogenic and benign classes. However, as increasing clinical classifications are deposited and shared publicly, for example via ClinVar, it is hoped

expanded truthsets can be constructed over time, and the performance of MAVEs re-quantified more robustly, and potentially at higher evidence strengths.

Overall, it is evident that this is a rapidly evolving field, and ongoing efforts are actively reshaping how MAVE data are standardised, calibrated, and incorporated into clinical variant interpretation. One such recent advance is the work by Zeiberg *et al.*[36], which addressed a key limitation in the clinical application of MAVE: the assignment of a single evidence strength to all variants exceeding an assay threshold. The authors introduced a gene-based calibration framework that converts continuous MAVE scores into variant-level ACMG/AMP PS3 and BS3 evidence strengths by estimating gene-specific pathogenicity priors and likelihood ratios. This approach replaces assay-wide thresholds with quantitative, variant-specific evidence assignment, enabling more accurate and tailored interpretation of functional data for clinical use. Complementing this, Calhoun *et al.*[37] proposed a structured framework for combining multiplexed functional data from multiple MAVE for the same gene, illustrated using four *TP53* assays, to generate integrated functional scores that increase PS3/BS3 evidence strength and reclassify additional VUS, further maximising the clinical utility of MAVE datasets. This is further supported by recent gene-specific calibration approaches, such as acmgscaler, which map functional scores directly onto ACMG/AMP evidence levels[38]. In parallel, ongoing international and national efforts led by ClinGen and initiatives such as CanVIG-UK are improving clinical translation through data sharing, standardisation and consensus interpretation, highlighting the increasing clinical utility of MAVE[29,39].

SHARING SGE DATA AND IMPROVING CLINICAL USABILITY OF MAVE RESOURCES

For SGE data to have clinical value, functional scores must be shared through stable and accessible resources rather than existing only within individual studies.

Recent guidance emphasises that clinical utility depends not only on assay quality but also on how variant effect data are released, as formalised in the guidelines proposed by Livesey *et al.*[40] for variant effect predictors (VEP) that are also relevant to MAVE. It is emphasized that variant effect data should be freely accessible, transparently generated, and accompanied by sufficient documentation to support independent evaluation and reuse. Although these guidelines were developed for computational tools, the same principles apply to functional datasets: scores should be reported relative to standardized reference sequences (e.g. MANE transcripts), use consistent variant nomenclature (e.g. HGVS), and include clear explanations of what the scores represent biologically.

Barriers to clinical use of functional data remain substantial. A survey of professionals involved in variant interpretation found that, although most clinicians use functional evidence, two-thirds reported that suitable functional data were rarely available when needed[41]. High-throughput functional assays were the category respondents felt least comfortable interpreting, despite their increasing prevalence[42]. The most common obstacles were limited access to standardized functional evidence, uncertainty regarding data quality, and lack of confidence in how results should be interpreted. Clinicians expressed strong demand for standardized interpretation, transparent performance metrics, and centralized access to primary data, particularly through resources linked to ClinVar.

To support this need, infrastructure for sharing MAVE data has expanded considerably. MaveDB[43] now serves as the primary community repository for multiplexed functional data and hosts over seven million variant effect measurements that are largely released under CC0 licence[44], enabling unrestricted reuse. In parallel, MaveRegistry[45] provides a coordination platform for the MAVE community by enabling target nomination, project registration and real-time progress tracking, helping to reduce duplication and promote collaboration across laboratories.

Together, these developments indicate that while technical barriers to sharing MAVE data are diminishing, clinical adoption depends on high-quality annotation and contextual metadata. In line with these principles, the SGE datasets generated in this thesis will be deposited in MaveDB and CanVIG-UK following publication and, where appropriate, linked to UKBB analyses to support transparency, reusability and clinical relevance. Broad sharing of calibrated functional data is therefore key to improving clinical interpretation of rare variants.

RESOLVING VUS USING MAVE DATA

Once MAVE data are interpreted within the ACMG/ClinGen framework, their main clinical value is the reduction of VUS[46]. Clinical and population-scale studies show that VUS account for roughly half of all reported variants and are more frequent among individuals of non-European genetic ancestry[46,47]. Unresolved VUS create uncertainty, unnecessary clinical interventions, and inconsistent communication of variant interpretation[48–51].

Gene-specific MAVE studies show that calibrated functional maps can resolve a substantial proportion of VUS. Integration of MAVE for *BRCA1*[52], *TP53*[46], *MSH2*[53], *DDX3X*[11], *BAP1*[19] and *RAD51C*[12] into ClinGen/ACMG frameworks led to reclassification of approximately 50-90% of variants. Beyond reclassification, Ranola *et al.*[54] showed that even when MAVE data do not substantially change ACMG classifications across a gene, they increase confidence by

providing evidence and reducing uncertainty across many variants. Thus, MAVEs improve not only classification outcomes, but also the certainty of overall evidence applied to all variants in the gene.

Extending these findings, the SGE data in this thesis distinguish variants with strong functional effects from those with no measurable impact and support reinterpretation alongside population frequency, tumour history and segregation evidence. In *POT1*, where over half of ClinVar coding variants were previously classified as VUS, SGE enables reassessment of nearly 90% of VUS, including rare missense and synonymous variants. In *TINF2*, functional profiling of cancer-relevant domains similarly identifies variants with likely LoF effects where clinical and functional evidence remains currently limited.

At a broader level, although single-gene studies such as those presented here demonstrate the effectiveness of MAVE, they have not directly addressed ancestry-related disparities in variant interpretation. This gap was addressed by Dawood *et al*.[46] who applied an ACMG/ClinGen-based pipeline incorporating calibrated MAVE data for *BRCA1*, *TP53* and *PTEN* to reclassify VUS from the All of Us and gnomAD[55] cohorts. Using standardised MAVE scores as PS3/BS3 evidence, rather than heterogeneous functional assays, led to higher reclassification rates for VUS from individuals of non-European-like ancestry and eliminated ancestry differences in unresolved VUS. In contrast to allele frequency and computational prediction criteria, which showed systematic bias, MAVE data operated equitably across populations, demonstrating that functional evidence can reduce ancestry-related uncertainty independently of population representation in reference databases.

These findings show that experimentally derived functional MAVE evidence improves VUS resolution and diagnostic utility and promotes more consistent variant interpretation across populations. In this light, this thesis envisions that the SGE data for *POT1* and *TINF2* may contribute to resolving uncertainty in the interpretation of rare germline variants and to supporting clinical decision-making across healthcare settings worldwide.

INTEGRATING SGE DATA INTO CLINICAL MANAGEMENT OF TELOMERE-ASSOCIATED CANCER

Beyond variant classification, SGE data have clear potential for clinical risk stratification and translational decision-making. In this thesis, SGE scores provide quantitative functional evidence that can support ACMG/AMP PS3/BS3 assignment and prioritise variants for further evaluation, when integrated with clinical and familial context, to inform genetic counselling and surveillance. Importantly, this utility extends beyond individual patients: functional characterisation can guide cascade testing within families and, support a targeted, phenotype-

driven testing strategy, as pathogenic *POT1* (and *TINF2*) variants are rare overall but enriched in clinically defined high-risk groups.

Nevertheless, functional evidence alone is insufficient for clinical decision-making, and SGE results should be interpreted within a broader clinical context. For *POT1*, data from this work support targeted dermatological surveillance in families at increased risk of melanoma, with consideration of associated malignancies, including haematological cancers and glioma[56,57]. Although *TINF2* is not yet suitable for routine inclusion in cancer gene panels due to the scarcity of validated pathogenic variants, SGE provides a forward-looking functional resource to support interpretation of newly identified variants and to help guiding future development of tumour surveillance guidelines.

SGE HAP1-A5 CELL MODEL LIMITATIONS

While this work provides a systematic functional analysis of variants in *POT1* and key regions of *TINF2*, several limitations should be considered when interpreting the results and assessing their clinical applicability.

As outlined in **Chapter 1**, *POT1* and *TINF2* CPLT variants are predominantly heterozygous and act mainly through partial LoF, with phenotypic consequences shaped primarily by the affected functional domain and cellular context, including replicative tissue demand and levels of telomerase expression. The haploid HAP1-A5 cell model evaluates each allele in isolation, in the absence of a wild-type allele. As a result, variants that impair telomere maintenance through partial protein insufficiency, impaired assembly of shelterin complexes, subtle defects in telomeric processing, or mechanisms that depend on allelic interactions may not reach the threshold required to produce measurable fitness effects in this assay, despite being clinically pathogenic[58]. Accordingly, the assay preferentially detects strong LoF effects, while gain-of-function (GoF), neomorphic, or dominant-negative variants may be underrepresented if they do not confer a measurable fitness disadvantage.

Some of these constraints are reflected in the observation that clinically established TBD variants[59,60] scored as functionally neutral/unchanged in the assay, indicating that certain pathogenic mechanisms underlying TBD are not captured by the HAP1-A5 fitness-based readout. Multiple features of the assay are likely to contribute to this.

First, HAP1-A5 cells have impaired classical non-homologous end joining (C-NHEJ) due to *LIG4* deletion[19], a modification required for efficient homology-directed repair (HDR) and thus enabling endogenous SGE. However suppression of C-NHEJ also biases DNA repair toward

alternative NHEJ (A-NHEJ) or microhomology-mediated end joining (MMEJ), via *LIG3*-associated processes[61]. As a consequence, telomere end-protection defects are more likely to result in error-prone fusion events, such that the assay preferentially detects variants that reach a viability-limiting threshold of telomere uncapping, rather than variants whose primary consequence is progressive telomere erosion[62].

Second, the 21-day screening window was selected to optimise sensitivity to strong functional defects while minimising spontaneous diploidisation but is likely insufficient for substantial telomere attrition to develop[63]. In addition, the near-haploid HAP1-A5 background may influence assay behaviour. HAP1 cells are derived from the KBM7 lineage and retain residual non-haploid genomic regions, and although haploid populations can be maintained, they are subject to ploidy instability during extended culture[64–66]. Importantly, apparent “diploidisation” reflects the selective outgrowth of higher-ploidy cells with a proliferative advantage rather than frequent genome duplication events, resulting in progressive karyotypic heterogeneity over time[67,68]. This raises the possibility that background genetic variation or structural alterations elsewhere in the genome could modify the observed effects of engineered *POT1* or *TINF2* variants. Although there is no clear evidence that haploidy per se imposes a distinct telomere-maintenance programme, altered gene dosage in a near-haploid context may shift the threshold at which telomere dysfunction translates into measurable fitness effects. In future applications, maintenance of haploidy could be further stabilised through approaches such as sorting via flow cytometric enrichment[63] or the use of small molecules (e.g. 10-deacetylbaccatin-III (10-DAB III) or Diclazuril) that selectively favour haploid cell populations[68], potentially reducing confounding effects of ploidy drift during screening.

Third, HAP1-A5 cells are immortalised and have very long telomeres, whereas individuals with TBD typically present with short telomeres prior to disease onset[69,70]. Because short-telomere syndromes are driven primarily by cumulative telomere erosion, rather than catastrophic end-joining, this difference in baseline telomere length reduces the probability that variants whose main effect is progressive telomere attrition will manifest as fitness defects within the assay window[71–73]. Fourth, HAP1-A5 cells do not recapitulate the cell-type–specific biology of tissues commonly affected in telomere disorders, such as melanocytes, glial cells, lung tissue or haematopoietic lineages. Tissue-specific factors such as telomerase activity, DDR signalling and chromatin state are likely to influence variant behaviour *in vivo* and may contribute to the disease-specific phenotypes observed in patients[74,75]. In addition, telomere-related variants may exert disease effects under conditions of replicative and metabolic stress, or inflammatory signalling, none of which are modelled or controlled in baseline culture. Overall, while SGE is highly effective in HAP1-A5 cells, modelling short-telomere TBDs (e.g. *DKC1*- or *RTEL1*-

associated dyskeratosis congenita) is likely to require diploid fibroblast-based systems such as WI-38 or IMR-90, including SV40- or E6/E7-immortalised derivatives, which could better recapitulate telomere attrition-driven pathology[68]. Future studies combining longer-term assays, lineage-specific models, together with controlled perturbations of cellular stress states, will be needed to capture these dimensions of telomere biology[76].

More generally, functional effects measured in a single model system cannot fully capture the clinical consequences of a variant. In patients, variant effects are likely to be shaped by broader genetic context, including differences in telomere gene expression, allelic dosage, additional germline or somatic variants in interacting pathways, and other epistatic effects that may modify penetrance or phenotype. As a result, some variants may behave differently in vivo than in the HAP1-A5 assay. This does not diminish the value of SGE but emphasises that these scores should be interpreted as context-defined functional evidence and integrated with orthogonal molecular and clinical data when used for clinical variant interpretation.

FUTURE DIRECTIONS IN SGE AND MAVE STUDIES

Ongoing technological and conceptual advances are rapidly expanding the scope and clinical relevance of MAVE approaches. Several priorities, therefore, emerge for extending the impact of SGE in telomere biology and related disease contexts.

MOVING BEYOND CELL FITNESS: RICHER PHENOTYPES FOR TELOMERE BIOLOGY GENES

In this thesis, variant effects were measured primarily through cellular fitness, leveraging the sensitivity of HAP1-A5 cells to acute telomere dysfunction. While this approach identified strong LoF variants, it could not resolve mechanistic subclasses of telomere dysfunction. Despite validation in **Chapter 4** showing telomere elongation and increased telomere-induced foci (TIFs) in HAP1 cells carrying deleterious *POT1* variants, broader phenotypic resolution is still required. As discussed in **Chapter 2**, MAVE can be increasingly extended beyond essentiality screens to include transcriptomic, imaging-based and multi-omic phenotypes. This will be particularly important for telomere genes, in which CPLT and TBD reflect distinct mechanisms affecting telomere length, end protection, replication stress, and DDR. Pooled SGE strategies could be extended to assay variants contributing to telomere damage, including TIF formation measured by multiplexed imaging of γH2AX/53BP1 and shelterin components[77,78]. Telomere length could be explored as a functional endpoint using flow cytometry with fluorescence *in situ* hybridization (FLOW-FISH)-based stratification, with

downstream analysis of variant distributions across telomere-length fractions[79], enabling detection of variants whose primary effects on telomere maintenance may not translate into acute fitness defects. Conceptually, this type of readout could be implemented in pooled MAVE formats by coupling quantitative telomere measurements to variant identity through sorting or imaging-based approaches. For example, recent optical pooled screening strategies combining telomeric DNA FISH with in situ sequencing demonstrate that telomere-associated fluorescence intensity can be measured at single-cell resolution and directly linked to specific genetic perturbations within a pooled population[80], enabled by the use of expressed guide RNA barcodes that are read out via in situ sequencing. In such frameworks, cells can be stratified based on telomeric signal intensity, serving as a proxy for telomere length or ALT activity, and variant enrichment across these strata quantified by sequencing[80]. In parallel, sequencing-based approaches such as long-read telomere profiling (e.g., Telo-seq) provide high-resolution measurements of telomere length at the level of individual chromosome arms and could, in principle, be integrated into future MAVE frameworks. However, this method currently lacks variant-level barcoding and has not yet been adapted for pooled single-cell screening, highlighting a trade-off between resolution and scalability[81]. Complementary to this, single-cell RNA sequencing could assess transcriptional stress responses and RNA-level defects, including splicing disruption and NMD, linking variants to intra cellular pathways and cellular states[18]. However, not all sequencing-based readouts are directly informative for telomere length. recent work[82] has shown that telomere-like reads in single-cell ATAC-seq do not reliably report telomere length, but instead largely reflect subtelomeric accessibility and broader chromatin condensation state. This suggests that ATAC-derived telomeric signals are more likely to serve as an orthogonal readout of chromatin state than as a direct pooled measure of telomere length.

Together, these advances suggest that imaging- or flow-based proxies of telomere length currently represent the most tractable route for incorporating telomere length as a quantitative phenotype in pooled MAVE designs. While substantial methodological development will be required to achieve the necessary throughput and robustness, these strategies provide a clear conceptual framework for extending MAVE beyond traditional fitness-based assays and towards richer, phenotype-resolved interrogation of telomere biology genes.

EXPANDING GENOMIC SCOPE BEYOND CODING REGIONS: CONTEXT- AND ISOFORM-AWARE FUNCTIONAL MAPPING

Although coding variants account for most clinically reported alterations in *POT1* and *TINF2*, they represent only a small fraction of the mutational landscape. The contribution of regulatory

variation, such as effects on gene expression, RNA processing, and transcript usage, remains largely unresolved[72,73].

Several experimental frameworks have now been established to functionally interrogate non-coding DNA at scale. CRISPR tiling[74,75] and CRISPR interference/activation (CRISPRi/a)[74–76] perturb endogenous regulatory elements, while massively parallel reporter assays (MPRAs)[77–82] quantify the transcriptional activity of thousands of sequences in parallel. These approaches have revealed functional variants in promoters[76], enhancers[76–78], silencers[79], untranslated regions (UTRs)[80,83] and splice junctions[81,84–86], showing that a subset of rare non-coding variants can exert substantial regulatory effects[87]. An example is the *TERT* promoter, where recurrent non-coding variants create *de novo* transcription-factor binding sites, enhancing *TERT* promoter activity and increasing transcription[88]. MPRAs confirmed the transcriptional consequences of these variants[89], highlighting that pathogenic processes in telomere biology can originate from regulatory rather than coding variation.

Beyond regulatory control, non-coding variants can also influence alternative splicing and transcript isoform usage, producing disease-relevant transcripts with distinct exon structures, UTRs, or regulatory profiles. Most functional screens target only a single canonical transcript, meaning that variants altering isoform abundance or splicing patterns may escape detection, as shown for certain cancer-susceptibility genes, where isoform context modifies pathogenicity[90,91]. Emerging transcript-resolved methods combining isoform-selective editing, long-read RNA sequencing, and quantitative transcriptomics[92,93], could improve the detection of clinically relevant splicing and expression effects. For telomere-maintenance genes, whose isoform expression varies across tissues and developmental stages[94,95], such approaches may refine genotype–phenotype interpretation.

Despite these advances, translating non-coding and isoform-level functional data into clinical variant interpretation remains an evolving goal. The primary challenge is the limited number of benign and pathogenic regulatory variant truthsets needed for assay calibration and benchmarking within clinical frameworks[96]. Furthermore, the regulatory impact of non-coding and isoform-dependent variants is highly context-specific, influenced by chromatin state, 3D genome organization, and cell-type-specific transcriptional programs. Therefore, future MAVEs must integrate both regulatory and transcript-level analyses and interpret their readouts within the endogenous chromatin and transcriptional context to achieve comprehensive variant interpretation[87,97].

FUNCTIONAL EXPANSION BEYOND HDR-BASED SGE

SGE represents one implementation within the broader MAVE framework. Complementary genome-editing approaches, including base and prime editing, expand the range of variant classes that can be interrogated and extend functional analysis beyond HDR-based designs. Base editing[109,110] enables high-throughput installation of single-nucleotide variants, while prime editing[111–114] allows access to variant classes that are difficult to interrogate by SGE, including small insertions, deletions and splice-altering mutations.

MAVE technologies have also been applied to diploid models, including large B-cell lymphoma lines (DLBCL)[115] and human mammary epithelial cells (HMECs)[116]. These assays could capture dominant-negative, GoF and hypomorphic variants in a heterozygous context, making the scores directly relevant for clinical interpretation in the corresponding tissues. However, applying MAVE in diploid systems also led to reduced editing efficiencies, mixed allele states, limited genomic coverage, and greater context dependence of functional readouts, highlighting not only the need for optimized editing and more informative phenotypic assays, but also the importance of model selection. An important consideration is that variant effects may not be fully conserved across cellular contexts. The original *BRCA1* SGE study[117] was performed in haploid HAP1 cells and established the power of endogenous editing for clinical variant classification, but more recent work by Dace et al.[116] suggests that a substantial subset of *BRCA1* variants show discordant behaviour between HAP1 and HMEC models, with many variants that impair function in HAP1 appearing neutral in mammary epithelial cells. Importantly, these discordant variants were shown to be hypomorphic and associated with intermediate cancer risk, rather than representing false-positive functional effects. These findings indicate that functional consequences can depend on lineage-specific biology, gene dosage, and the particular cellular dependencies of the model system, including compensatory DNA repair pathways and alternative splicing mechanisms present in specific cell types, and therefore support the inclusion of disease-relevant cell types when interpreting variants in cancer predisposition genes. More broadly, comparison across cell types can refine variant classification by distinguishing high-penetrance loss-of-function alleles from variants conferring reduced or context-dependent risk. For telomere biology genes, this is likely to be especially important, given the marked differences in telomerase activity, replication stress, chromatin state, and proliferative demand across tissues implicated in melanoma, glioma, haematological malignancy, and pulmonary fibrosis.

Moreover, recent work by Fayer *et al.*[118] developed an induced pluripotent stem cell (iPSC)-based SGE framework that edited over 1,500 variants into *MYBPC3* and *POLG* and measured their effects in cardiomyocytes, neurons and cardiac organoids. By extending saturation-scale

endogenous editing beyond monolayer cultures, this study demonstrates that MAVE can be performed in systems that partially recapitulate tissue organisation and multicellular context. Together, these findings establish the feasibility of interrogating variant effects in disease-relevant cellular environments, providing a foundation for analogous studies of telomere biology genes in lineages implicated in melanoma, haematological malignancy and glioma[104,118].

MACHINE LEARNING AND MAVE

As multiple MAVE datasets accumulate for the same gene, it is increasingly clear that no single assay captures all disease mechanisms[104]. This limitation is well illustrated by studies of *TP53*, where saturation mutagenesis across multiple cellular contexts showed that while variants with severe functional consequences are consistently detected, many missense variants exhibit subtle or context-dependent effects that are not resolved by any single assay[119]. To address this, model-based integration approaches have been developed that combine multiple independent MAVE datasets into a unified functional score per variant using machine learning (ML) classifiers[119,120]. Such integrated scores improved discrimination of pathogenic and benign ClinVar variants and enhanced application of ACMG/ClinGen functional evidence codes.

Beyond gene-specific integration, ML frameworks increasingly leverage large-scale sequence features, population frequency data, and available MAVE resources to enable scalable prediction of variant effects across many genes[109–111]. In this context, experimentally derived MAVE data serve as essential anchors for model training, validation, and calibration, ensuring that predictions remain grounded in empirical functional evidence rather than purely *in silico* inference.

Despite this increasing integration, divergence between MAVE-derived functional scores and machine learning-based variant effect predictors is still observed. As discussed in Chapter 1, *in silico* tools such as AlphaMissense[121,122] and related predictors[123] are scalable and informative[124], but they remain indirect models based largely on sequence conservation and structural features and may disagree with one another or overestimate pathogenicity for variants with limited functional impact. By contrast, SGE measures variant consequences experimentally at the endogenous locus and can therefore detect effects on cellular fitness, splicing, and transcript regulation that are not fully captured by protein-level prediction alone. As a result, discordance may arise for variants with context-dependent or mechanism-specific effects, or where predictors over- or under-estimate functional impact[125–128]. Importantly, such divergence is not necessarily indicative of error in either approach but highlights the

complementary nature of experimental and computational methods and provides a basis for prioritising variants for further functional investigation. ML can leverage this by using existing MAVE data to guide the prioritisation of future functional assays. By training models on MAVE-derived scores together with sequence features and genomic annotations, variants and genomic regions, including non-coding elements, can be prioritised according to their likelihood of yielding informative functional effects. This enables a more efficient and hypothesis-driven deployment of saturation mutagenesis, reducing experimental scale and cost while retaining empirical functional assays as the primary source of evidence.

CONCLUSION

This thesis demonstrates that SGE is a practical and scalable approach for the functional interpretation of rare variants in telomere-associated cancer genes, generating high-resolution endogenous variant effect maps for *POT1* and *TINF2* that capture functional consequences at nucleotide resolution.

SGE of *POT1* revealed that variant effects span a continuum of functional impact rather than conforming to a binary pathogenic-benign model. The degree of functional depletion correlated quantitatively with telomere elongation and cancer risk in population-based and familial datasets. Importantly, this work represents the first functional genomics screen to directly link quantitative cellular phenotypes to variant penetrance at the population level, integrating SGE-derived functional severity with UKBB data, familial cancer cohorts and pedigrees, as well as telomere length and telomeric DDR signalling. Together, these analyses provide empirical support for a proportional model of genetic risk in telomere-associated cancer predisposition.

SGE of *TINF2* identified functionally constrained protein-protein interaction surfaces within the shelterin complex and provided experimental support for the pathogenicity of both established and newly identified variants. The work presented here refined the tumour spectrum associated with *TINF2* disruption and demonstrates the value of functional evidence in genes where clinical data is sparse, and variant interpretation remains challenging.

A central contribution of this work is the integration of SGE data with established clinical variant interpretation frameworks. For *POT1*, functional scores were quantitatively calibrated to ACMG/AMP criteria, providing PS3/BS3-weighted evidence and demonstrating that SGE can meet key requirements for clinically relevant functional assays when appropriate truthsets are available. Across both assayed genes, SGE data reduced the burden of VUS by supplying

objective functional evidence in settings where clinical and population data alone are insufficient.

At the same time, this work highlights important constraints of the HAP1-A5 haploid fitness-based model, which is sensitive to cancer-predisposition mechanisms and does not capture the full spectrum of TBDs. These limitations underscore the need to view SGE as one component of a broader functional MAVE toolkit, complemented by alternative model systems, phenotypic readouts, and assay designs.

Overall, this thesis establishes experimentally grounded functional evidence for the interpretation of genetic variation in *POT1* and *TINF2*. While further work extending these methodologies and analytical frameworks to additional variant classes, cellular models, and clinical cohorts will be required, the findings presented here demonstrate that high-throughput functional genomics can substantially reduce diagnostic uncertainty and strengthen the clinical interpretation of rare variants in hereditary cancer genetics. In doing so, this work positions SGE as a practical translational bridge between genomic discovery and clinical decision-making in telomere-associated cancer syndromes.

REFERENCES

1. Jaber, D., Zhang, J., and Godley, L.A. (2025). Detecting likely germline variants during tumor-based molecular profiling. J. Clin. Invest. *135*, e190264. https://doi.org/10.1172/JCI190264.
2. Burren, O.S., Dhindsa, R.S., Deevi, S.V.V., Wen, S., Nag, A., Mitchell, J., Hu, F., Loesch, D.P., Smith, K.R., Razdan, N., et al. (2024). Genetic architecture of telomere length in 462,666 UK Biobank whole-genome sequences. Nat. Genet. *56*, 1832–1840. https://doi.org/10.1038/s41588-024-01884-7.
3. Chen, E., Facio, F.M., Aradhya, K.W., Rojahn, S., Hatchell, K.E., Aguilar, S., Ouyang, K., Saitta, S., Hanson-Kwan, A.K., Capurro, N.N., et al. (2023). Rates and Classification of Variants of Uncertain Significance in Hereditary Disease Genetic Testing. JAMA Netw. Open *6*, e2339571. https://doi.org/10.1001/jamanetworkopen.2023.39571.
4. Rehm, H.L., Alaimo, J.T., Aradhya, S., Bayrak-Toydemir, P., Best, H., Brandon, R., Buchan, J.G., Chao, E.C., Chen, E., Clifford, J., et al. (2023). The landscape of reported VUS in multi-gene panel and genomic testing: Time for a change. Genet. Med. *25*, 100947. https://doi.org/10.1016/j.gim.2023.100947.
5. LaDuca, H., Polley, E.C., Yussuf, A., Hoang, L., Gutierrez, S., Hart, S.N., Yadav, S., Hu, C., Na, J., Goldgar, D.E., et al. (2020). A clinical guide to hereditary cancer panel testing: evaluation of gene-specific cancer associations and sensitivity of genetic testing criteria in a cohort of 165,000 high-risk patients. Genet. Med. *22*, 407–415. https://doi.org/10.1038/s41436-019-0633-8.
6. Savage, S.A., Bertuch, A.A., and Team Telomere and the Clinical Care Consortium for Telomere-Associated Ailments (CCCTAA) (2025). Different phenotypes with different endings—Telomere biology disorders and cancer predisposition with long telomeres. Br. J. Haematol. *206*, 69–73. https://doi.org/10.1111/bjh.19851.
7. Hart, T., Tong, A.H.Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., Chandrashekhar, M., Hustedt, N., Seth, S., Noonan, A., et al. (2017). Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. G3 GenesGenomesGenetics *7*, 2719–2727. https://doi.org/10.1534/g3.117.041277.
8. Behan, F.M., Iorio, F., Picco, G., Gonçalves, E., Beaver, C.M., Migliardi, G., Santos, R., Rao, Y., Sassi, F., Pinnelli, M., et al. (2019). Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature *568*, 511–516. https://doi.org/10.1038/s41586-019-1103-9.

9. Baumann, P., and Cech, T.R. (2001). Pot1, the Putative Telomere End-Binding Protein in Fission Yeast and Humans. Science *292*, 1171–1175. https://doi.org/10.1126/science.1060036.
10. Calvete, O., Garcia-Pavia, P., Domínguez, F., Bougeard, G., Kunze, K., Braeuninger, A., Teule, A., Lasa, A., Ramón Y Cajal, T., Llort, G., et al. (2017). The wide spectrum of POT1 gene variants correlates with multiple cancer types. Eur. J. Hum. Genet. EJHG *25*, 1278–1281. https://doi.org/10.1038/ejhg.2017.134.
11. Radford, E.J., Tan, H.-K., Andersson, M.H.L., Stephenson, J.D., Gardner, E.J., Ironfield, H., Waters, A.J., Gitterman, D., Lindsay, S., Abascal, F., et al. (2023). Saturation genome editing of DDX3X clarifies pathogenicity of germline and somatic variation. Nat. Commun. *14*, 7702. https://doi.org/10.1038/s41467-023-43041-4.
12. Olvera-León, R., Zhang, F., Offord, V., Zhao, Y., Tan, H.K., Gupta, P., Pal, T., Robles-Espinoza, C.D., Arriaga-González, F.G., Matsuyama, L.S.A.S., et al. (2024). High-resolution functional mapping of RAD51C by saturation genome editing. Cell *187*, 5719-5734.e19. https://doi.org/10.1016/j.cell.2024.08.039.
13. De Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev. *19*, 2100–2110. https://doi.org/10.1101/gad.1346005.
14. De Lange, T. (2018). Shelterin-Mediated Telomere Protection. Annu. Rev. Genet. *52*, 223–247. https://doi.org/10.1146/annurev-genet-032918-021921.
15. Xu, M., Kiselar, J., Whited, T.L., Hernandez-Sanchez, W., and Taylor, D.J. (2019). POT1-TPP1 differentially regulates telomerase via POT1 His266 and as a function of single-stranded telomere DNA length. Proc. Natl. Acad. Sci. U. S. A. *116*, 23527–23533. https://doi.org/10.1073/pnas.1905381116.
16. Sekne, Z., Ghanim, G.E., Van Roon, A.-M.M., and Nguyen, T.H.D. (2022). Structural basis of human telomerase recruitment by TPP1-POT1. Science *375*, 1173–1176. https://doi.org/10.1126/science.abn6840.
17. Baumann, P., and Price, C. (2010). Pot1 and telomere maintenance. FEBS Lett. *584*, 3779–3784. https://doi.org/10.1016/j.febslet.2010.05.024.
18. Buckley, M., Kajba, C.M., Forrester, N., Terwagne, C., Sawyer, C., Shepherd, S.T.C., Jonghe, J.D., Dace, P., Turajlic, S., and Findlay, G.M. (2023). Saturation Genome Editing Resolves the Functional Spectrum of Pathogenic *VHL* Alleles (Genomics) https://doi.org/10.1101/2023.06.10.542698.
19. Waters, A.J., Brendler-Spaeth, T., Smith, D., Offord, V., Tan, H.K., Zhao, Y., Obolenski, S., Nielsen, M., Van Doorn, R., Murphy, J.-E., et al. (2024). Saturation genome editing of BAP1 functionally classifies somatic and germline variants. Nat. Genet. *56*, 1434–1445. https://doi.org/10.1038/s41588-024-01799-3.
20. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. Cell *176*, 535-548.e24. https://doi.org/10.1016/j.cell.2018.12.015.
21. Schmutz, I., Mensenkamp, A.R., Takai, K.K., Haadsma, M., Spruijt, L., De Voer, R.M., Choo, S.S., Lorbeer, F.K., Van Grinsven, E.J., Hockemeyer, D., et al. (2020). TINF2 is a haploinsufficient tumor suppressor that limits telomere length. eLife *9*, e61235. https://doi.org/10.7554/eLife.61235.
22. Jensen, M.R., Jelsig, A.M., Gerdes, A.-M., Hölmich, L.R., Kainu, K.H., Lorentzen, H.F., Hansen, M.H., Bak, M., Johansson, P.A., Hayward, N.K., et al. (2023). TINF2 is a major susceptibility gene in Danish patients with multiple primary melanoma. Hum. Genet. Genomics Adv. *4*, 100225. https://doi.org/10.1016/j.xhgg.2023.100225.
23. He, H., Li, W., Comiskey, D.F., Liyanarachchi, S., Nieminen, T.T., Wang, Y., DeLap, K.E., Brock, P., and de la Chapelle, A. (2020). A Truncating Germline Mutation of TINF2 in Individuals with Thyroid Cancer or Melanoma Results in Longer Telomeres. Thyroid Off. J. Am. Thyroid Assoc. *30*, 204–213. https://doi.org/10.1089/thy.2019.0156.
24. DeBoy, E.A., Nicosia, A.M., Liyanarachchi, S., Iyer, S.S., Shah, M.H., Ringel, M.D., Brock, P., and Armanios, M. (2024). Telomere-lengthening germline variants predispose to a syndromic papillary thyroid cancer subtype. Am. J. Hum. Genet. *111*, 1114–1124. https://doi.org/10.1016/j.ajhg.2024.04.006.
25. Ballinger, M.L., Pattnaik, S., Mundra, P.A., Zaheed, M., Rath, E., Priestley, P., Baber, J., Ray-Coquard, I., Isambert, N., Causeret, S., et al. (2023). Heritable defects in telomere and mitotic function selectively predispose to sarcomas. Science *379*, 253–260. https://doi.org/10.1126/science.abj4784.
26. Koivuluoma, S., Vorimo, S., Mattila, T.M., Tervasmäki, A., Kumpula, T., Kuismin, O., Winqvist, R., Moilanen, J., Mantere, T., and Pylkäs, K. (2023). Truncating TINF2 p.Tyr312Ter variant and inherited breast cancer susceptibility. Fam. Cancer *22*, 13–17. https://doi.org/10.1007/s10689-022-00295-z.
27. Liu, D., O'Connor, M.S., Qin, J., and Songyang, Z. (2004). Telosome, a Mammalian Telomere-associated Complex Formed by Multiple Telomeric Proteins. J. Biol. Chem. *279*, 51338–51342. https://doi.org/10.1074/jbc.M409293200.

28. On behalf of the Brotman Baty Institute Mutational Scanning Working Group, Gelman, H., Dines, J.N., Berg, J., Berger, A.H., Brnich, S., Hisama, F.M., James, R.G., Rubin, A.F., Shendure, J., et al. (2019). Recommendations for the collection and use of multiplexed functional data for clinical variant interpretation. Genome Med. *11*, 85. https://doi.org/10.1186/s13073-019-0698-7.
29. Allen, S., Garrett, A., Muffley, L., Fayer, S., Foreman, J., Adams, D.J., Hurles, M., Rubin, A.F., Roth, F.P., Starita, L.M., et al. (2024). Workshop report: the clinical application of data from multiplex assays of variant effect (MAVEs), 12 July 2023. Eur. J. Hum. Genet. EJHG *32*, 593–600. https://doi.org/10.1038/s41431-024-01566-2.
30. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). 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. Genet. Med. *17*, 405–424. https://doi.org/10.1038/gim.2015.30.
31. Sequence Variant Interpretation Working Group. SVI general recommendations for using ACMG/AMP criteria.
32. Brnich, S.E., Abou Tayoun, A.N., Couch, F.J., Cutting, G.R., Greenblatt, M.S., Heinen, C.D., Kanavy, D.M., Luo, X., McNulty, S.M., Starita, L.M., et al. (2019). Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med. *12*, 3. https://doi.org/10.1186/s13073-019-0690-2.
33. ENIGMA ClinGen ENIGMA BRCA1 and BRCA2 Expert Panel Specifications to the ACMG/AMP Variant Interpretation Guidelines for BRCA1 Version 1.0.0 2023.
34. Fortuno, C., Lee, K., Olivier, M., Pesaran, T., Mai, P.L., Andrade, K.C., Attardi, L.D., Crowley, S., Evans, D.G., Feng, B., et al. (2021). Specifications of the ACMG/AMP variant interpretation guidelines for germline *TP53* variants. Hum. Mutat. *42*, 223–236. https://doi.org/10.1002/humu.24152.
35. Mester, J.L., Ghosh, R., Pesaran, T., Huether, R., Karam, R., Hruska, K.S., Costa, H.A., Lachlan, K., Ngeow, J., Barnholtz-Sloan, J., et al. (2018). Gene-specific criteria for *PTEN* variant curation: Recommendations from the ClinGen PTEN Expert Panel. Hum. Mutat. *39*, 1581–1592. https://doi.org/10.1002/humu.23636.
36. Zeiberg, D., Tejura, M., McEwen, A.E., Fayer, S., Pejaver, V., Rubin, A.F., Starita, L.M., Fowler, D.M., O'Donnell-Luria, A., and Radivojac, P. (2025). Gene-based calibration of high-throughput functional assays for clinical variant classification. Preprint, https://doi.org/10.1101/2025.04.29.651326 https://doi.org/10.1101/2025.04.29.651326.
37. Calhoun, J.D., Dawood, M., Rowlands, C.F., Fayer, S., Radford, E.J., McEwen, A.E., Turnbull, C., Spurdle, A.B., Starita, L.M., and Jagannathan, S. (2025). Combining multiplexed functional data to improve variant classification. ArXiv, arXiv:2503.18810v1.
38. Badonyi, M., and Marsh, J.A. (2025). acmgscaler : an R package and Colab for standardized gene-level variant effect score calibration within the ACMG/AMP framework. Bioinformatics *41*, btaf503. https://doi.org/10.1093/bioinformatics/btaf503.
39. Allen, S., Loong, L., Garrett, A., Torr, B., Durkie, M., Drummond, J., Callaway, A., Robinson, R., Burghel, G.J., Hanson, H., et al. (2024). Recommendations for laboratory workflow that better support centralised amalgamation of genomic variant data: findings from CanVIG-UK national molecular laboratory survey. J. Med. Genet. *61*, 305–312. https://doi.org/10.1136/jmg-2023-109645.
40. Livesey, B.J., Badonyi, M., Dias, M., Frazer, J., Kumar, S., Lindorff-Larsen, K., McCandlish, D.M., Orenbuch, R., Shearer, C.A., Muffley, L., et al. (2025). Guidelines for releasing a variant effect predictor. Genome Biol. *26*, 97. https://doi.org/10.1186/s13059-025-03572-z.
41. Park, M.S., Kumar, R.D., Ovadiuc, C., Folta, A., McEwen, A.E., Snyder, A., Fowler, D.M., Rubin, A.F., Shirts, B.H., Starita, L.M., et al. (2025). Insights on improving accessibility and usability of functional data to unlock its potential for variant interpretation. Preprint, https://doi.org/10.1101/2025.01.25.25321117 https://doi.org/10.1101/2025.01.25.25321117.
42. Fowler, D.M., Adams, D.J., Gloyn, A.L., Hahn, W.C., Marks, D.S., Muffley, L.A., Neal, J.T., Roth, F.P., Rubin, A.F., Starita, L.M., et al. (2023). An Atlas of Variant Effects to understand the genome at nucleotide resolution. Genome Biol. *24*, 147, s13059-023-02986–x. https://doi.org/10.1186/s13059-023-02986-x.
43. Rubin, A.F., Stone, J., Bianchi, A.H., Capodanno, B.J., Da, E.Y., Dias, M., Esposito, D., Frazer, J., Fu, Y., Grindstaff, S.B., et al. (2025). MaveDB 2024: a curated community database with over seven million variant effects from multiplexed functional assays. Genome Biol. *26*, 13. https://doi.org/10.1186/s13059-025-03476-y.
44. Creative Commons — CC0 1.0 Universal CC0 1.0 Universal.

45. Kuang, D., Weile, J., Kishore, N., Nguyen, M., Rubin, A.F., Fields, S., Fowler, D.M., and Roth, F.P. (2021). MaveRegistry: a collaboration platform for multiplexed assays of variant effect. Bioinforma. Oxf. Engl. *37*, 3382–3383. https://doi.org/10.1093/bioinformatics/btab215.
46. Dawood, M., Fayer, S., Pendyala, S., Post, M., Kalra, D., Patterson, K., Venner, E., Muffley, L.A., Fowler, D.M., Rubin, A.F., et al. (2024). Using multiplexed functional data to reduce variant classification inequities in underrepresented populations. Genome Med. *16*, 143. https://doi.org/10.1186/s13073-024-01392-7.
47. Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. *42*, D980–D985. https://doi.org/10.1093/nar/gkt1113.
48. Gould, D., Walker, R., Makari-Judson, G., and Seven, M. (2022). Experiences of individuals with a variant of uncertain significance on genetic testing for hereditary cancer risks: a mixed method systematic review. J. Community Genet. *13*, 371–379. https://doi.org/10.1007/s12687-022-00600-4.
49. Welsh, J.L., Hoskin, T.L., Day, C.N., Thomas, A.S., Cogswell, J.A., Couch, F.J., and Boughey, J.C. (2017). Clinical Decision-Making in Patients with Variant of Uncertain Significance in BRCA1 or BRCA2 Genes. Ann. Surg. Oncol. *24*, 3067–3072. https://doi.org/10.1245/s10434-017-5959-3.
50. Culver, J., Brinkerhoff, C., Clague, J., Yang, K., Singh, K., Sand, S., and Weitzel, J. (2013). Variants of uncertain significance in *BRCA* testing: evaluation of surgical decisions, risk perception, and cancer distress. Clin. Genet. *84*, 464–472. https://doi.org/10.1111/cge.12097.
51. Folta, A., Sedeño Cortés, A.E., Gupta, P., McEwen, A.E., Kao, E.Y., Horike-Pyne, M., Stone, J., Shirts, B.H., Dubard-Gault, M.E., Fowler, D.M., et al. (2025). Imprecision medicine: Systematic gaps in reporting variants of uncertain significance (VUS) and their reclassifications. Genet. Med. *27*, 101501. https://doi.org/10.1016/j.gim.2025.101501.
52. Fayer, S., Horton, C., Dines, J.N., Rubin, A.F., Richardson, M.E., McGoldrick, K., Hernandez, F., Pesaran, T., Karam, R., Shirts, B.H., et al. (2021). Closing the gap: Systematic integration of multiplexed functional data resolves variants of uncertain significance in BRCA1, TP53, and PTEN. Am. J. Hum. Genet. *108*, 2248–2258. https://doi.org/10.1016/j.ajhg.2021.11.001.
53. Scott, A., Hernandez, F., Chamberlin, A., Smith, C., Karam, R., and Kitzman, J.O. (2022). Saturation-scale functional evidence supports clinical variant interpretation in Lynch syndrome. Genome Biol. *23*, 266. https://doi.org/10.1186/s13059-022-02839-z.
54. Ranola, J.M.O., Horton, C., Pesaran, T., Fayer, S., Starita, L.M., and Shirts, B.H. (2024). Assigning credit where it is due: an information content score to capture the clinical value of multiplexed assays of variant effect. BMC Bioinformatics *25*, 295. https://doi.org/10.1186/s12859-024-05920-5.
55. Gudmundsson, S., Singer-Berk, M., Watts, N.A., Phu, W., Goodrich, J.K., Solomonson, M., Genome Aggregation Database Consortium, Rehm, H.L., MacArthur, D.G., and O'Donnell-Luria, A. (2022). Variant interpretation using population databases: Lessons from gnomAD. Hum. Mutat. *43*, 1012–1030. https://doi.org/10.1002/humu.24309.
56. Accardo, M.-L., Osborne, J., and Else, T. (1993). POT1 Tumor Predisposition. In GeneReviews®, M. P. Adam, S. Bick, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, and A. Amemiya, eds. (University of Washington, Seattle).
57. Tsoulaki, O., Evans, D.G., Sinha, K., Rajan, N., Bakr, F., Hatcher, H., Napolitano, A., Finn, E., Iyengar, S., Sohaib, A., et al. (2025). UK clinical practice guidelines for the management of patients with constitutional *POT1* pathogenic variants. J. Med. Genet. *62*, 559–565. https://doi.org/10.1136/jmg-2025-110638.
58. Llargués-Sistac, G., Bonjoch, L., and Castellvi-Bel, S. (2023). HAP1, a new revolutionary cell model for gene editing using CRISPR-Cas9. Front. Cell Dev. Biol. *11*, 1111488. https://doi.org/10.3389/fcell.2023.1111488.
59. Kelich, J., Aramburu, T., Van Der Vis, J.J., Showe, L., Kossenkov, A., Van Der Smagt, J., Massink, M., Schoemaker, A., Hennekam, E., Veltkamp, M., et al. (2022). Telomere dysfunction implicates POT1 in patients with idiopathic pulmonary fibrosis. J. Exp. Med. *219*, e20211681. https://doi.org/10.1084/jem.20211681.
60. Takai, H., Jenkinson, E., Kabir, S., Babul-Hirji, R., Najm-Tehrani, N., Chitayat, D.A., Crow, Y.J., and De Lange, T. (2016). A POT1 mutation implicates defective telomere end fill-in and telomere truncations in Coats plus. Genes Dev. *30*, 812–826. https://doi.org/10.1101/gad.276873.115.
61. Simsek, D., Brunet, E., Wong, S.Y.-W., Katyal, S., Gao, Y., McKinnon, P.J., Lou, J., Zhang, L., Li, J., Rebar, E.J., et al. (2011). DNA Ligase III Promotes Alternative Nonhomologous End-Joining during Chromosomal Translocation Formation. PLoS Genet. *7*, e1002080. https://doi.org/10.1371/journal.pgen.1002080.

62. Jones, R.E., Oh, S., Grimstead, J.W., Zimbric, J., Roger, L., Heppel, N.H., Ashelford, K.E., Liddiard, K., Hendrickson, E.A., and Baird, D.M. (2014). Escape from Telomere-Driven Crisis Is DNA Ligase III Dependent. Cell Rep. *8*, 1063–1076. https://doi.org/10.1016/j.celrep.2014.07.007.
63. Beigl, T.B., Kjosås, I., Seljeseth, E., Glomnes, N., and Aksnes, H. (2020). Efficient and crucial quality control of HAP1 cell ploidy status. Biol. Open *9*, bio057174. https://doi.org/10.1242/bio.057174.
64. Andersson, B.S., Collins, V.P., Kurzrock, R., Larkin, D.W., Childs, C., Ost, A., Cork, A., Trujillo, J.M., Freireich, E.J., and Siciliano, M.J. (1995). KBM-7, a human myeloid leukemia cell line with double Philadelphia chromosomes lacking normal c-ABL and BCR transcripts. Leukemia *9*, 2100–2108.
65. Essletzbichler, P., Konopka, T., Santoro, F., Chen, D., Gapp, B.V., Kralovics, R., Brummelkamp, T.R., Nijman, S.M.B., and Bürckstümmer, T. (2014). Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. Genome Res. *24*, 2059–2065. https://doi.org/10.1101/gr.177220.114.
66. Banerjee, R., Sotero-Caio, C.G., Fu, B., and Yang, F. (2022). Chromosomal instability (CIN) in HAP1 cell lines revealed by multiplex fluorescence in situ hybridisation (M-FISH). Mol. Cytogenet. *15*, 46. https://doi.org/10.1186/s13039-022-00625-x.
67. Olbrich, T., Mayor-Ruiz, C., Vega-Sendino, M., Gomez, C., Ortega, S., Ruiz, S., and Fernandez-Capetillo, O. (2017). A p53-dependent response limits the viability of mammalian haploid cells. Proc. Natl. Acad. Sci. *114*, 9367–9372. https://doi.org/10.1073/pnas.1705133114.
68. Olbrich, T., Vega-Sendino, M., Murga, M., de Carcer, G., Malumbres, M., Ortega, S., Ruiz, S., and Fernandez-Capetillo, O. (2019). A Chemical Screen Identifies Compounds Capable of Selecting for Haploidy in Mammalian Cells. Cell Rep *28*, 597-604.e4. https://doi.org/10.1016/j.celrep.2019.06.060.
69. Alter, B.P., Baerlocher, G.M., Savage, S.A., Chanock, S.J., Weksler, B.B., Willner, J.P., Peters, J.A., Giri, N., and Lansdorp, P.M. (2007). Very short telomere length by flow fluorescence in situ hybridization identifies patients with dyskeratosis congenita. Blood *110*, 1439–1447. https://doi.org/10.1182/blood-2007-02-075598.
70. Savage, S.A. (2022). Dyskeratosis congenita and telomere biology disorders. Hematology *2022*, 637–648. https://doi.org/10.1182/hematology.2022000394.
71. Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. Nature *345*, 458–460. https://doi.org/10.1038/345458a0.
72. Counter, C.M., Avilion, A.A., LeFeuvre, C.E., Stewart, N.G., Greider, C.W., Harley, C.B., and Bacchetti, S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. *11*, 1921–1929. https://doi.org/10.1002/j.1460-2075.1992.tb05245.x.
73. Murnane, J.P. (2012). Telomere dysfunction and chromosome instability. Mutat. Res. - Fundam. Mol. Mech. Mutagen. *730*, 28–36. https://doi.org/10.1016/j.mrfmmm.2011.04.008.
74. Batista, L.F.Z., Dokal, I., and Parker, R. (2022). Telomere biology disorders: time for moving towards the clinic? Trends Mol. Med. *28*, 882–891. https://doi.org/10.1016/j.molmed.2022.08.001.
75. Kam, M.L.W., Nguyen, T.T.T., and Ngeow, J.Y.Y. (2021). Telomere biology disorders. Npj Genomic Med. *6*, 36. https://doi.org/10.1038/s41525-021-00198-5.
76. Sinnott-Armstrong, N., Fields, S., Roth, F., Starita, L.M., Trapnell, C., Villen, J., Fowler, D.M., and Queitsch, C. (2024). Understanding genetic variants in context. eLife *13*, e88231. https://doi.org/10.7554/eLife.88231.
77. Choudhary, K., and McManus, M.T. (2025). Scalable imaging-based profiling of CRISPR perturbations with protein barcodes. Preprint, https://doi.org/10.1101/2025.11.11.687767 https://doi.org/10.1101/2025.11.11.687767.
78. Azeroglu, B., Khurana, S., Wang, S.-C., Tricola, G.M., Sharma, S., Jubelin, C., Cortolezzis, Y., Pegoraro, G., Miller, K.M., Stracker, T.H., et al. (2025). Identification of modulators of the ALT pathway through a native FISH-based optical screen. Cell Rep. *44*, 115114. https://doi.org/10.1016/j.celrep.2024.115114.
79. Freen-van Heeren, J.J. (2024). Combining CRISPR with Flow-FISH to study CRISPR -mediated genome perturbation. Cytometry A *105*, 7–9. https://doi.org/10.1002/cyto.a.24815.
80. Quintanilla, I., Azeroglu, B., Sagar, M.A.K., Stracker, T.H., Denchi, E.L., and Pegoraro, G. (2025). Optical pooled screening for the discovery of regulators of the alternative lengthening of telomeres pathway. Methods *241*, 1–12. https://doi.org/10.1016/j.ymeth.2025.05.001.
81. Schmidt, T.T., Tyer, C., Rughani, P., Haggblom, C., Jones, J.R., Dai, X., Frazer, K.A., Gage, F.H., Juul, S., Hickey, S., et al. (2024). High resolution long-read telomere sequencing reveals dynamic mechanisms in aging and cancer. Nat. Commun. *15*, 5149. https://doi.org/10.1038/s41467-024-48917-7.

82. Yakovenko, I., Mihai, I.S., Selinger, M., Rosenbaum, W., Dernstedt, A., Gröning, R., Trygg, J., Carroll, L., Forsell, M., and Henriksson, J. (2025). Telomemore enables single-cell analysis of cell cycle and chromatin condensation. Nucleic Acids Res. *53*, gkaf031. https://doi.org/10.1093/nar/gkaf031.
83. Bourdon, J.-C., Fernandes, K., Murray-Zmijewski, F., Liu, G., Diot, A., Xirodimas, D.P., Saville, M.K., and Lane, D.P. (2005). p53 isoforms can regulate p53 transcriptional activity. Genes Dev. *19*, 2122–2137. https://doi.org/10.1101/gad.1339905.
84. Tammaro, C., Raponi, M., Wilson, D.I., and Baralle, D. (2012). *BRCA1* exon 11 alternative splicing, multiple functions and the association with cancer. Biochem. Soc. Trans. *40*, 768–772. https://doi.org/10.1042/BST20120140.
85. Réal, A., Brown, A., Yung, G.P., Borel, C., Lykoskoufis, N., Seebach, J., Dermitzakis, E., Ramisch, A., and Viñuela, A. (2024). Direct long-read RNA sequencing uncovers functional variation affecting transcript production and RNA modifications. Preprint, https://doi.org/10.21203/rs.3.rs-4613444/v1 https://doi.org/10.21203/rs.3.rs-4613444/v1.
86. Chen, Y., Davidson, N.M., Wan, Y.K., Yao, F., Su, Y., Gamaarachchi, H., Sim, A., Patel, H., Low, H.M., Hendra, C., et al. (2025). A systematic benchmark of Nanopore long-read RNA sequencing for transcript-level analysis in human cell lines. Nat. Methods *22*, 801–812. https://doi.org/10.1038/s41592-025-02623-4.
87. Subasri, M., Shooshtari, P., Watson, A.J., and Betts, D.H. (2021). Analysis of TERT Isoforms across TCGA, GTEx and CCLE Datasets. Cancers *13*, 1853. https://doi.org/10.3390/cancers13081853.
88. Kim, J.J., Ahn, A., Ying, J.Y., and Ludlow, A.T. (2025). Discovery and characterization of a novel telomerase alternative splicing isoform that protects lung cancer cells from chemotherapy induced cell death. Sci. Rep. *15*, 6787. https://doi.org/10.1038/s41598-025-90639-3.
89. French, J.D., and Edwards, S.L. (2020). The Role of Noncoding Variants in Heritable Disease. Trends Genet. *36*, 880–891. https://doi.org/10.1016/j.tig.2020.07.004.
90. Zhang, F., and Lupski, J.R. (2015). Non-coding genetic variants in human disease: Figure 1. Hum. Mol. Genet. *24*, R102–R110. https://doi.org/10.1093/hmg/ddv259.
91. Canver, M.C., Bauer, D.E., and Orkin, S.H. (2017). Functional interrogation of non-coding DNA through CRISPR genome editing. Methods *121–122*, 118–129. https://doi.org/10.1016/j.ymeth.2017.03.008.
92. Montalbano, A., Canver, M.C., and Sanjana, N.E. (2017). High-Throughput Approaches to Pinpoint Function within the Noncoding Genome. Mol. Cell *68*, 44–59. https://doi.org/10.1016/j.molcel.2017.09.017.
93. Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer–promoter connections with CRISPR interference. Science *354*, 769–773. https://doi.org/10.1126/science.aag2445.
94. Kheradpour, P., Ernst, J., Melnikov, A., Rogov, P., Wang, L., Zhang, X., Alston, J., Mikkelsen, T.S., and Kellis, M. (2013). Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. Genome Res. *23*, 800–811. https://doi.org/10.1101/gr.144899.112.
95. Birnbaum, R.Y., Patwardhan, R.P., Kim, M.J., Findlay, G.M., Martin, B., Zhao, J., Bell, R.J.A., Smith, R.P., Ku, A.A., Shendure, J., et al. (2014). Systematic Dissection of Coding Exons at Single Nucleotide Resolution Supports an Additional Role in Cell-Specific Transcriptional Regulation. PLoS Genet. *10*, e1004592. https://doi.org/10.1371/journal.pgen.1004592.
96. Doni Jayavelu, N., Jajodia, A., Mishra, A., and Hawkins, R.D. (2020). Candidate silencer elements for the human and mouse genomes. Nat. Commun. *11*, 1061. https://doi.org/10.1038/s41467-020-14853-5.
97. Sample, P.J., Wang, B., Reid, D.W., Presnyak, V., McFadyen, I.J., Morris, D.R., and Seelig, G. (2019). Human 5′ UTR design and variant effect prediction from a massively parallel translation assay. Nat. Biotechnol. *37*, 803–809. https://doi.org/10.1038/s41587-019-0164-5.
98. Julien, P., Miñana, B., Baeza-Centurion, P., Valcárcel, J., and Lehner, B. (2016). The complete local genotype–phenotype landscape for the alternative splicing of a human exon. Nat. Commun. *7*, 11558. https://doi.org/10.1038/ncomms11558.
99. Shukla, C.J., McCorkindale, A.L., Gerhardinger, C., Korthauer, K.D., Cabili, M.N., Shechner, D.M., Irizarry, R.A., Maass, P.G., and Rinn, J.L. (2018). High-throughput identification of RNA nuclear enrichment sequences. EMBO J. *37*, e98452. https://doi.org/10.15252/embj.201798452.
100. Oikonomou, P., Goodarzi, H., and Tavazoie, S. (2014). Systematic Identification of Regulatory Elements in Conserved 3′ UTRs of Human Transcripts. Cell Rep. *7*, 281–292. https://doi.org/10.1016/j.celrep.2014.03.001.

101. Mueller, W.F., Larsen, L.S.Z., Garibaldi, A., Hatfield, G.W., and Hertel, K.J. (2015). The Silent Sway of Splicing by Synonymous Substitutions. J. Biol. Chem. *290*, 27700–27711. https://doi.org/10.1074/jbc.M115.684035.
102. Ke, S., Anquetil, V., Zamalloa, J.R., Maity, A., Yang, A., Arias, M.A., Kalachikov, S., Russo, J.J., Ju, J., and Chasin, L.A. (2018). Saturation mutagenesis reveals manifold determinants of exon definition. Genome Res. *28*, 11–24. https://doi.org/10.1101/gr.219683.116.
103. Chong, R., Insigne, K.D., Yao, D., Burghard, C.P., Wang, J., Hsiao, Y.-H.E., Jones, E.M., Goodman, D.B., Xiao, X., and Kosuri, S. (2019). A Multiplexed Assay for Exon Recognition Reveals that an Unappreciated Fraction of Rare Genetic Variants Cause Large-Effect Splicing Disruptions. Mol. Cell *73*, 183-194.e8. https://doi.org/10.1016/j.molcel.2018.10.037.
104. McEwen, A.E., Tejura, M., Fayer, S., Starita, L.M., and Fowler, D.M. (2025). Multiplexed assays of variant effect for clinical variant interpretation. Nat. Rev. Genet. https://doi.org/10.1038/s41576-025-00870-x.
105. Ellingford, J.M., Ahn, J.W., Bagnall, R.D., Baralle, D., Barton, S., Campbell, C., Downes, K., Ellard, S., Duff-Farrier, C., FitzPatrick, D.R., et al. (2022). Recommendations for clinical interpretation of variants found in non-coding regions of the genome. Genome Med. *14*, 73. https://doi.org/10.1186/s13073-022-01073-3.
106. Cao, C., Gong, W., Shuai, Y., Rasouli, S., Ge, Q., Khan, A., Dakic, A., Putluri, N., Shvets, G., Zheng, Y.-L., et al. (2025). Canonical and non-canonical functions of the non-coding RNA component (TERC) of telomerase complex. Cell Biosci. *15*, 30. https://doi.org/10.1186/s13578-025-01367-0.
107. Florez-Vargas, O., Ho, M., Hogshead, M.H., Papenberg, B.W., Lee, C.-H., Forsythe, K., Jones, K., Luo, W., Teshome, K., Blauwendraat, C., et al. (2025). Genetic regulation of TERT splicing affects cancer risk by altering cellular longevity and replicative potential. Nat. Commun. *16*, 1676. https://doi.org/10.1038/s41467-025-56947-y.
108. Kircher, M., Xiong, C., Martin, B., Schubach, M., Inoue, F., Bell, R.J.A., Costello, J.F., Shendure, J., and Ahituv, N. (2019). Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. Nat. Commun. *10*, 3583. https://doi.org/10.1038/s41467-019-11526-w.
109. Coelho, M.A., Cooper, S., Strauss, M.E., Karakoc, E., Bhosle, S., Gonçalves, E., Picco, G., Burgold, T., Cattaneo, C.M., Veninga, V., et al. (2023). Base editing screens map mutations affecting interferon-γ signaling in cancer. Cancer Cell *41*, 288-303.e6. https://doi.org/10.1016/j.ccell.2022.12.009.
110. Coelho, M.A., Strauss, M.E., Watterson, A., Cooper, S., Bhosle, S., Illuzzi, G., Karakoc, E., Dinçer, C., Vieira, S.F., Sharma, M., et al. (2024). Base editing screens define the genetic landscape of cancer drug resistance mechanisms. Nat. Genet. *56*, 2479–2492. https://doi.org/10.1038/s41588-024-01948-8.
111. Herger, M., Kajba, C.M., Buckley, M., Cunha, A., Strom, M., and Findlay, G.M. (2025). High-throughput screening of human genetic variants by pooled prime editing. Cell Genomics *5*, 100814. https://doi.org/10.1016/j.xgen.2025.100814.
112. Reyes, M.A., Herger, M., Cubitt, L., and Findlay, G.M. (2025). PETRA: Prime editing of transcribed regulatory elements to assay expression. Preprint, https://doi.org/10.1101/2025.11.18.689114 https://doi.org/10.1101/2025.11.18.689114.
113. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature *576*, 149–157. https://doi.org/10.1038/s41586-019-1711-4.
114. Erwood, S., Bily, T.M.I., Lequyer, J., Yan, J., Gulati, N., Brewer, R.A., Zhou, L., Pelletier, L., Ivakine, E.A., and Cohn, R.D. (2022). Saturation variant interpretation using CRISPR prime editing. Nat. Biotechnol. *40*, 885–895. https://doi.org/10.1038/s41587-021-01201-1.
115. Meitlis, I., Allenspach, E.J., Bauman, B.M., Phan, I.Q., Dabbah, G., Schmitt, E.G., Camp, N.D., Torgerson, T.R., Nickerson, D.A., Bamshad, M.J., et al. (2020). Multiplexed Functional Assessment of Genetic Variants in CARD11. Am. J. Hum. Genet. *107*, 1029–1043. https://doi.org/10.1016/j.ajhg.2020.10.015.
116. Dace, P., Forrester, N.M., Zanti, M., Cubitt, L., Terwagne, C., Buckley, M., Van Veen, E.M., Wilson, T.S., Scaffidi, P., Michailidou, K., et al. (2025). Saturation genome editing of *BRCA1* across cell types accurately resolves cancer risk. Preprint, https://doi.org/10.1101/2025.08.11.25333423 https://doi.org/10.1101/2025.08.11.25333423.
117. Findlay, G.M., Daza, R.M., Martin, B., Zhang, M.D., Leith, A.P., Gasperini, M., Janizek, J.D., Huang, X., Starita, L.M., and Shendure, J. (2018). Accurate classification of BRCA1 variants with saturation genome editing. Nature *562*, 217–222. https://doi.org/10.1038/s41586-018-0461-z.
118. Fayer, S., Garge, R.K., Hopkins, M., Friedman, C.E., McGee, A.V., Rico, J., Powell, R.L., McDermot, E., Smith, N.T., Pendyala, S., et al. (2025). Editing stem cell genomes at scale to measure variant effects in

diverse cell and genetic contexts. Preprint, https://doi.org/10.1101/2025.11.12.25340127 https://doi.org/10.1101/2025.11.12.25340127.
119. Giacomelli, A.O., Yang, X., Lintner, R.E., McFarland, J.M., Duby, M., Kim, J., Howard, T.P., Takeda, D.Y., Ly, S.H., Kim, E., et al. (2018). Mutational processes shape the landscape of TP53 mutations in human cancer. Nat. Genet. *50*, 1381–1387. https://doi.org/10.1038/s41588-018-0204-y.
120. Boettcher, S., Miller, P.G., Sharma, R., McConkey, M., Leventhal, M., Krivtsov, A.V., Giacomelli, A.O., Wong, W., Kim, J., Chao, S., et al. (2019). A dominant-negative effect drives selection of *TP53* missense mutations in myeloid malignancies. Science *365*, 599–604. https://doi.org/10.1126/science.aax3649.
121. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
122. Ljungdahl, A., Kohani, S., Page, N.F., Wells, E.S., Wigdor, E.M., Dong, S., and Sanders, S.J. (2023). AlphaMissense is better correlated with functional assays of missense impact than earlier prediction algorithms (Genetics) https://doi.org/10.1101/2023.10.24.562294.
123. Frazer, J., Notin, P., Dias, M., Gomez, A., Min, J.K., Brock, K., Gal, Y., and Marks, D.S. (2021). Disease variant prediction with deep generative models of evolutionary data. Nature *599*, 91–95. https://doi.org/10.1038/s41586-021-04043-8.
124. Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L.H., Zielinski, M., Sargeant, T., et al. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science *381*, eadg7492. https://doi.org/10.1126/science.adg7492.
125. Molotkov, I., Mardis, E.R., and Artomov, M. (2024). Making sense of missense: challenges and opportunities in variant pathogenicity prediction. Dis. Model. Mech. *17*, dmm052218. https://doi.org/10.1242/dmm.052218.
126. Kong, S.W., Lee, I.-H., Collen, L.V., Field, M., Manrai, A.K., Snapper, S.B., and Mandl, K.D. (2025). Discordance between a deep learning model and clinical-grade variant pathogenicity classification in a rare disease cohort. Npj Genom Med *10*, 17. https://doi.org/10.1038/s41525-025-00480-w.
127. Grimm, D.G., Azencott, C., Aicheler, F., Gieraths, U., MacArthur, D.G., Samocha, K.E., Cooper, D.N., Stenson, P.D., Daly, M.J., Smoller, J.W., et al. (2015). The Evaluation of Tools Used to Predict the Impact of Missense Variants Is Hindered by Two Types of Circularity. Hum. Mutat. *36*, 513–523. https://doi.org/10.1002/humu.22768.
128. Starita, L.M., Ahituv, N., Dunham, M.J., Kitzman, J.O., Roth, F.P., Seelig, G., Shendure, J., and Fowler, D.M. (2017). Variant Interpretation: Functional Assays to the Rescue. Am. J. Hum. Genet. *101*, 315–325. https://doi.org/10.1016/j.ajhg.2017.07.014.