Mutation Signatures and Genome Protection: Applications of WES in Chemoprevention Trials

Pravin Badhe

Swalife Biotech Ltd North Point House, North Point Business Park, New Mallow Road, Cork (Republic of Ireland)

Corresponding author: drpravinbadhe@swalifebiotech.com
Doi: 10.5281/zenodo.18734285
Received: 14 January 2026
Accepted: 21 January 2026

Abstract

Whole-exome sequencing (WES) has emerged as a transformative technology for understanding cancer initiation and progression through the lens of mutational signatures characteristic patterns of somatic mutations imprinted on cancer genomes by underlying mutagenic processes. This review synthesizes current knowledge on the application of WES-derived mutation signatures as biomarkers and endpoints in cancer chemoprevention trials, with emphasis on genomic approaches to preventing malignant transformation. We examine mechanistic underpinnings of major carcinogen-specific signatures (tobacco, ultraviolet, alcohol, aging), discuss technical considerations for implementing WES as a clinical platform, and outline strategies for integrating mutation signature analysis into adaptive trial designs. Particular focus is given to oral squamous cell carcinoma (OSCC) prevention, where loss of heterozygosity (LOH) profiling and emerging somatic mutation burden metrics offer unprecedented opportunities for patient enrichment and molecular monitoring. We review the paradigm shift from histologic endpoints (intraepithelial neoplasia) to genomic endpoints (tumor mutational burden, signature composition, clonal evolution), and discuss how mutation signatures can guide personalized chemoprevention through synthetic lethal approaches targeting DNA damage response deficiencies. The integration of multi-omic biomarkers, liquid biopsy, and adaptive trial designs promises to accelerate the clinical translation of genome protection strategies, ultimately improving cancer prevention outcomes in high-risk populations.

Keywords: mutation signatures, whole-exome sequencing, cancer chemoprevention, WES biomarkers, genomic instability, DNA damage, oral cancer prevention, carcinogen exposure

1. Introduction:

Cancer develops through the stepwise accumulation of somatic mutations in key oncogenes and tumor suppressor genes, a process that typically spans years to decades and offers multiple intervention windows.[1] The traditional cancer prevention paradigm focused on primary prevention (eliminating carcinogen exposure) and secondary prevention (detecting early-stage disease). However, emerging evidence demonstrates that intermediate stages of transformation characterized by specific patterns of genomic alterations in histologically normal or premalignant epithelium represent actionable targets for chemoprevention. The advent of next-generation sequencing (NGS) has fundamentally altered our ability to detect, characterize, and monitor these molecular lesions in real time, enabling mechanistically informed prevention strategies.[2]

Mutational signatures, defined as characteristic patterns of somatic mutations reflecting the underlying etiology of DNA damage and repair processes, provide a molecular "fingerprint" of carcinogenic exposures and cellular vulnerabilities.[3] Unlike individual driver mutations (TP53, KRAS, CDKN2A), which may be present in both normal and malignant cells with variable penetrance, mutational signatures capture the composite mutagenic

burden imposed by exogenous (tobacco, ultraviolet radiation, chemical exposures) and endogenous (aging, inflammation, replication stress) processes. Whole-exome sequencing (WES), which sequences the entire protein-coding exome at high coverage, enables comprehensive characterization of these signatures with sufficient statistical power to extract individual components from complex multidimensional datasets.[4]

This review examines the convergence of mutational signature science and cancer chemoprevention, with particular emphasis on how WES-based genomic profiling can advance the field beyond histologic and biochemical endpoints toward mechanistically grounded, mutation-specific prevention strategies. We highlight oral squamous cell carcinoma (OSCC) as a paradigmatic system for WES-guided chemoprevention, discuss standardization and clinical implementation challenges, and outline a framework for integrating genome protection into next-generation cancer prevention trials.

2. Mutational Signatures: Biological Mechanisms and Classification

2.1 Definition and Conceptual Framework

Mutational signatures are, in essence, the "scars" left by mutagenic processes on the cancer genome.[3] Each process whether viral infection, chemical carcinogen exposure, or intrinsic cellular dysfunction imprints a characteristic pattern of mutations determined by (1) the type of DNA damage incurred, (2) the DNA repair or replication pathways engaged, and (3) the sequence context flanking mutated bases. Critically, signatures distinguish between historical (past) mutational processes, which inform public health strategies but are no longer actively driving tumorigenesis, and ongoing processes, which represent potential therapeutic targets.[3]

Mathematical decomposition of somatic mutations using non-negative matrix factorization (NMF) and related algorithms extracts individual signatures from pooled datasets by treating the problem as a "blind source separation" challenge. The human cancer genome contains approximately 6 × 10^9 base pairs; given this enormous dimensionality, the power to detect signatures requires large mutation burdens and statistical sophistication. Whole-exome sequencing, which typically identifies 1,000–10,000 somatic single-nucleotide variants (SNVs) per tumor, provides sufficient coverage for reliable signature extraction in most solid tumors.[4]

2.2 Classification of Mutational Signatures

Signatures are classified by mutation class: single-base substitutions (SBS), double-base substitutions (DBS), small insertions and deletions (indels), and structural variations.[3] Each class reflects distinct mutagenic mechanisms and DNA repair processes.

Single-Base Substitution (SBS) Signatures: SBS signatures are characterized in 96-trinucleotide format, considering the immediate 5' and 3' flanking bases (e.g., TpCpG, where the C is mutated to T). The COSMIC database currently catalogs >70 distinct SBS signatures in human cancers, each linked to specific etiologies.[5] The most prevalent SBS signatures include:

- SBS1/SBS1b (Aging): C·G→T·A transitions at CpG dinucleotides due to spontaneous 5-methylcytosine deamination. Occurs at a rate of ~100 bp per telomere per cell division and correlates strongly with patient age at cancer diagnosis across 25+ cancer types.[3]

- SBS4 (Tobacco): Characterized by C·G→T·A transversions with transcriptional strand bias, found in 33.6% of head and neck cancers and accounting for 6.3% of mean mutational burden in tobacco-exposed tumors.[6] SBS4 signature arises from benzo[a]pyrene and other polycyclic aromatic hydrocarbons in tobacco smoke.

- SBS92 (Tobacco/Bladder): T·C→T transitions found predominantly in bladder cancer (7.6% of HNC samples), representing a distinct tobacco-related mutagenic mechanism, possibly related to tobacco nitrosamines NNN and NNK.[6,7]

- SBS7 (UV Radiation): Characterized by C·G→T·A mutations at dipyrimidines (CC→TT transitions up to 25% of mutation burden), strongly associated with cutaneous melanoma and squamous cell carcinoma, reflecting transcription-coupled nucleotide excision repair of UV-induced pyrimidine dimers.[3]

- SBS2/SBS13 (APOBEC): TpC-context mutations due to activation-induced cytidine deaminase (AICDA) and apolipoprotein B mRNA editing enzyme (APOBEC) activity, found in breast cancer and multiple other malignancies, with strand-coordinated mutations indicating single-stranded DNA targeting.[3]

- SBS16 (Alcohol): Alcohol-related signature with characteristic mutation patterns, co-occurring with tobacco signatures in esophageal and other malignancies, suggesting synergistic mutagenesis.[6]

Indel and Structural Variation Signatures: Signature 3 (microhomology-mediated indels) associates with BRCA1/BRCA2 deficiency, reflecting impaired homologous recombination repair and reliance on error-prone non-homologous end joining (NHEJ).[3] Signature 6 (microsatellite instability, MSI) reflects mismatch repair (MMR) deficiency and manifests as 1–3 bp indels within repetitive sequences, characteristic of Lynch syndrome and sporadic MMR-deficient colorectal cancers.[3]

2.3 Mechanistic Basis of Carcinogen-Specific Signatures

The tobacco-related SBS4 signature exemplifies how signature analysis reveals mechanistic insights relevant to chemoprevention. Tobacco smoke contains >60 carcinogens, yet SBS4 predominantly reflects benzo[a]pyrene (BaP) metabolism. BaP undergoes hepatic activation to BaP diol epoxide, which forms bulky DNA adducts; these lesions trigger transcription-coupled nucleotide excision repair (TCR), which explains the transcriptional strand bias observed in SBS4 (fewer mutations on the actively transcribed strand than on the non-transcribed strand).[3] Critically, SBS4 mutations are enriched in early, clonal events during head and neck cancer development, suggesting that tobacco-induced mutagenesis occurs in normal or premalignant epithelium years before invasive transformation.[6]

Recent WES studies in 265 head and neck cancer samples identified not only SBS4 and SBS92 but also a novel tobacco-associated signature, SBS_I, characterized by T·A→A·T transversions with transcriptional strand bias, potentially reflecting adenine damage from other tobacco carcinogens or metabolites like acrylamide.[6] The anatomical distribution of tobacco signatures varies considerably: SBS4 burden increases from oral cavity (17.3%) to oropharynx (17.4%) to larynx (66.7%), reflecting differences in carcinogen exposure patterns and epithelial vulnerability.[6] This anatomical heterogeneity has direct implications for chemoprevention trial design and patient stratification.

3. Whole-Exome Sequencing as a Genomic Platform for Prevention Trials

3.1 Technical Foundations and Quality Metrics

WES involves sequencing the ~50 Mb of protein-coding exons to high coverage (typically 50-500×), followed by alignment to a reference genome, variant calling, and annotation.[4] Compared to panel sequencing (targeting a fixed set of 50-1,000+ genes) and whole-genome sequencing (WGS, 3 billion base pairs), WES offers distinct advantages and tradeoffs:

Advantages: (1) Comprehensive coverage of the exome enables unbiased detection of driver and passenger mutations, (2) Complex biomarkers (tumor mutational burden, homologous recombination deficiency, microsatellite instability) can be accurately measured rather than estimated, (3) Data can be re-analyzed indefinitely as new targets or signatures emerge, and (4) Standardization across laboratories is more feasible than with custom-designed panels.[4]

Limitations: (1) Intronic and regulatory variants are missed (with some exceptions), (2) Copy number alterations are detected but at lower resolution than CNV-specific methods, and (3) Costs and turnaround time (TAT) remain barriers to routine clinical implementation, though rapidly declining.[4]

A multicenter pilot study involving five German institutions demonstrated high concordance of WES across non-harmonized protocols: tumor mutational burden (TMB) showed Pearson R correlation of 0.97-0.99 between institutions, somatic single-nucleotide variants had 91-95% positive percentage agreement, and complex biomarkers (HRD, MSI) showed strong concordance.[8] This study provides evidence that WES can be implemented as a clinical-grade platform with appropriate quality assurance.

3.2 Mutation Burden Quantification and Signature Attribution

Tumor mutational burden (TMB), quantified as mutations per megabase of sequenced exome, is a validated biomarker associated with response to immune checkpoint inhibitors (ICIs) in MMR-deficient and POLE-mutant cancers.[9] WES enables TMB calculation with superior accuracy compared to panel sequencing, as the denominator (Mb sequenced) and numerator (somatic mutations called) are both measured rather than estimated.[8]

Attribution of individual signatures to a patient's tumor requires deconvolving the composite mutational landscape. Computational tools such as mmsig (optimized for WES data) fit the observed 96-trinucleotide spectrum to a known signature catalog (COSMIC), yielding an estimated percentage contribution of each signature to the total mutation burden. For example, in a tobacco-exposed patient with oral dysplasia, WES + signature analysis might reveal 40% SBS4 (tobacco-specific), 30% SBS5 (aging), and 30% SBS1b (aging), indicating that approximately 40% of the mutation burden is attributable to ongoing tobacco exposure a quantifiable target for cessation and chemoprevention.[6]

4. Molecular Endpoints in Cancer Chemoprevention: Evolution from Histology to Genomics

4.1 Traditional Phase II Endpoints: Intraepithelial Neoplasia and Biomarkers

Cancer prevention trials traditionally rely on intermediate endpoints because clinical efficacy trials (with invasive cancer as the primary outcome) require enormous sample sizes and years of follow-up. The National Cancer Institute (NCI) Division of Cancer Prevention has sponsored hundreds of Phase I/II chemoprevention trials, which have shaped the field's understanding of feasible and informative endpoints.[10]

Histologic Endpoints: Intraepithelial neoplasia (IEN), graded as mild, moderate, or severe dysplasia, represents a histologically defined intermediate in the progression to invasive malignancy. Oral leukoplakia, characterized by oral epithelial dysplasia, has a meta-analysis-derived progression rate of 12.1% to invasive oral cancer over a mean of 4.3 years.[9,10] However, IEN endpoints are fraught with limitations: lesions show spontaneous regression (up to 30-46% of lesions regress without intervention), sampling bias from biopsy-induced healing,

and high natural variability. The EPOC trial of erlotinib in oral dysplasia enrolled 1,040 individuals to identify 562 candidates for bronchoscopy, ultimately randomizing only 150 LOH-positive patients illustrating the enormous screening burden imposed by IEN-based designs.[11]

Cancer-Associated Biomarkers: Ki-67 proliferation index, measured by immunohistochemistry in biopsied tissue, has been used as a surrogate for chemopreventive efficacy. The phase II trial of topical 4-hydroxytamoxifen in ductal carcinoma in situ of the breast showed that Ki-67 reduction correlated with improved disease-free survival when the comparison agent (tamoxifen) was known to be effective but Ki-67 modulation in isolation provides limited insight.[10]

Gene Expression Signatures: A paradigm shift is underway toward molecular signatures extracted from normal epithelium. Gustafson and colleagues identified a PI3K pathway activation signature in normal bronchial epithelium of smokers with lung cancer, which was reversible following myo-inositol treatment suggesting that pathway-specific genomic signatures in at-risk but histologically normal tissue might identify optimal candidate cohorts and endpoints.[10] This approach is more scalable than tissue biopsy-dependent endpoints and may better reflect the biology driving progression.

4.2 Emerging Genomic Endpoints: Loss of Heterozygosity, Tumor Mutational Burden, and Clonal Evolution

Loss of Heterozygosity (LOH) as a Genomic Biomarker: The Erlotinib Prevention of Oral Cancer (EPOC) trial represented a watershed moment in precision cancer prevention. This randomized, double-blind, placebo-controlled trial enrolled 379 patients with oral premalignant lesions (OPLs), stratified by LOH profiling at chromosomal loci associated with oral cancer risk (3p14, 9p21, and others).[11] LOH-positive patients (defined by loss at ≥1 critical locus) comprised 254 individuals, of whom 150 were randomized to erlotinib (150 mg/day) or placebo for 12 months.

Key findings were highly informative for chemoprevention strategy: (1) LOH positivity was validated as a prognostic marker of oral cancer risk 3-year cancer-free survival (CFS) was 74% for LOH+ patients versus 87% for LOH- patients (HR 2.19, p=0.01), (2) LOH-positive status correlated with increased EGFR gene copy number ($p<0.001$) and reduced CFS (p=0.01), and (3) Erlotinib did not improve CFS (70% erlotinib vs. 74% placebo, HR 1.27, p=0.45), despite being the rational molecular target.[11] However, erlotinib-induced skin rash (grade ≥2) paradoxically predicted improved CFS, suggesting off-target or immune-mediated mechanisms.

The EPOC trial's negative efficacy result underscores a fundamental principle: even validated biomarkers for risk stratification do not automatically predict therapeutic benefit. Nevertheless, LOH profiling has been incorporated into routine clinical practice for OPL risk assessment, and the EPOC paradigm (biomarker-selected, precision medicine prevention trial with molecular endpoints) has become a template for future studies.

Tumor Mutational Burden (TMB) as a Prevention Endpoint: TMB, typically measured as nonsynonymous mutations per megabase, is increasingly recognized as relevant to chemoprevention. In normal or premalignant epithelium exposed to carcinogens, TMB elevation reflects ongoing mutagenic exposure and predicts future cancer risk. For instance, patients with Lynch syndrome (MMR-deficient) accumulate somatic mutations at ~100× the rate of MMR-proficient individuals; this hypermutator phenotype stratifies Lynch carriers at the highest risk for colorectal cancer, and certain chemopreventive strategies (e.g., aspirin) are specifically targeted to this population.[12]

WES enables measurement of TMB with high accuracy, making it a feasible endpoint for chemoprevention trials. Adaptive trial designs could implement real-time TMB monitoring: if a chemopreventive agent fails to reduce TMB trajectories over 6-12 months, patients could be switched to alternative interventions or intensified surveillance. Such real-time molecular monitoring has been demonstrated with circulating tumor DNA (ctDNA) in therapeutic oncology settings and is increasingly feasible for prevention applications.[13]

Mutation Signature Burden and Composition: Beyond total TMB, the specific composition of mutational signatures that is, the percentage contribution of SBS4 (tobacco) versus SBS5 (aging) offers mechanistically richer information. A tobacco-exposed patient with premalignant oral dysplasia showing 50% SBS4 burden represents a distinct clinical entity from one with 10% SBS4 (and 90% aging-related SBS5). Chemopreventive interventions could be matched to signature composition: a patient with high SBS4 might benefit from tobacco-specific DNA repair enhancement, whereas one with high SBS5 and no smoking history might warrant different strategies.[14]

Recent advances in signature-guided therapeutics include synthetic lethal approaches. For example, patients with MMR deficiency (associated with Signature 6, microsatellite instability) harbor frameshift neoantigens and are candidated for immunoprevention with neoantigen-targeted vaccines in Lynch syndrome carriers.[14] Similarly, PARP inhibitors are being explored in chemoprevention settings for individuals with BRCA1/2 germline mutations (Signature 3: microhomology-mediated indels).[15]

5. WES Applications in Oral Squamous Cell Carcinoma: A Paradigmatic System

5.1 Molecular Landscape of OSCC and Tobacco-Related Signatures

OSCC represents approximately 3% of all cancers globally but carries disproportionate morbidity due to late-stage presentation, aggressive biology, and poor prognosis (5-year overall survival <50%).[16] Tobacco and alcohol use are the primary risk factors, followed by HPV infection and oral lichen planus. Recent large-scale WES and RNA-seq studies have mapped the OSCC genomic landscape, revealing recurrent driver mutations (TP53, FAT1, CDKN2A, KMT2C) and distinct mutational signatures tied to exposures.[17]

Tobacco-exposed OSCC demonstrates high SBS4 signature burden. A study of 265 head and neck cancer samples found SBS4 in 33.6% of HNC cases (accounting for 6.3% of mean TMB) and a novel SBS_I signature in tobacco-exposed patients.[6] The anatomical distribution is striking: SBS4 appears in 66.7% of laryngeal cancers versus only 17.3% of oral cavity cancers, reflecting differential carcinogen deposition and epithelial susceptibility.[6] These anatomical differences directly inform chemoprevention strategy: high-SBS4 laryngeal lesions might warrant more aggressive intervention than low-SBS4 oral cavity lesions.

Co-exposure to tobacco and alcohol synergistically increases mutagenesis. Alcohol-related SBS16 signature frequently co-occurs with tobacco signatures, with combined effects on mutation burden and driver gene patterns. For example, co-exposed patients show increased C>A driver mutations (24.9% smokers vs. 17.3% non-smokers, p=0.038), consistent with SBS4 dominance, but additionally accumulate alcohol-related T>A and other patterns.[6]

5.2 Preclinical Models and Translational Insights

The 4-nitroquinoline-1-oxide (4NQO)-induced oral cancer model provides a powerful preclinical system for understanding tobacco-induced carcinogenesis and testing chemoprevention strategies. 4NQO is a DNA adduct-forming agent that mimics tobacco-induced carcinogenesis by generating TP53 mutations (particularly hot spots G245 and R248) and resembles tobacco smoke exposure through immune microenvironment modulation.[18]

Critically, WES analysis of 4NQO-induced murine oral squamous cell carcinoma (4MOSC) cell lines revealed that 4NQO-induced mutation patterns exhibit 94% similarity to human tobacco signature SBS4, compared to only 40% similarity for DMBA (dimethylbenz[a]anthracene), a commonly used tobacco carcinogen.[18] This finding

validates 4NQO as a superior model for tobacco-related OSCC and provides a framework for testing chemoprevention agents in mutation-signature-matched preclinical systems.

The 4MOSC model has demonstrated complex immune microenvironment alterations: 4NQO exposure induces B and T cell populations' significant decrease in the spleen early in precancerous lesion development, with immune derangement persisting through malignant transformation.[19] This early immune suppression suggests that immunoprevention strategies such as checkpoint inhibitor therapy or neoantigen-targeted vaccines might be effective in preventing progression from dysplasia to invasive carcinoma.

6. Genome Protection Strategies: From Mechanism to Chemoprevention

6.1 Five Priority Targets Against Genomic Instability

Recent comprehensive reviews have identified five priority targets for preventing genomic instability and cancer development:[20,21]

1. Prevention of DNA Damage: Avoiding exogenous carcinogens (tobacco cessation, sun protection, occupational hazard mitigation) remains the gold standard but is often insufficient in high-risk populations with historical exposures or genetic susceptibilities.

2. Enhancement of DNA Repair: Upregulation of endogenous repair mechanisms through dietary and pharmacologic interventions (e.g., nucleotide excision repair, base excision repair, homologous recombination pathways).

3. Targeting Deficient DNA Repair: Exploiting synthetic lethal vulnerabilities in individuals with germline or early somatic DNA damage response (DDR) defects (e.g., PARP inhibitors in BRCA-mutant populations).

4. Impairing Centrosome Clustering: Cancer cells frequently harbor centrosome amplification and overcome lethality through clustering mechanisms; disrupting clustering selectively targets tumor cells while sparing normal cells.[22]

5. Inhibition of Telomerase Activity: Telomerase is reactivated in the vast majority of cancers; inhibition could theoretically prevent indefinite replicative potential during premalignant progression.

6.2 Nutraceutical and Dietary Interventions

Evidence supports several dietary components in reducing genomic instability and cancer risk:

- Vitamins B and D: Nucleotide metabolism and epigenetic regulation; nicotinamide (vitamin B3) demonstrated efficacy in reducing nonmelanoma skin cancer and actinic keratosis risk in a randomized trial (NNT ~23).[23]

- Selenium and Selenocysteine: Induce reactive oxygen species formation that preferentially kills DNA repair-deficient cancer cells without harming normal fibroblasts; selenium levels correlate inversely with cancer risk in prospective cohort studies.[20]

- Carotenoids, Polyphenols, Isothiocyanates: Multiple mechanisms including PARP activation, CYP1A inhibition, and Nrf2 pathway upregulation.[20,21]

- Resveratrol: Found in grapes and red wine; activates sirtuins and enhances DNA repair capacity.[20]

Clinical translation has been limited by small sample sizes and heterogeneous biomarker endpoints. The Lung Cancer Chemoprevention Study of leucoselect phytosome (grape seed extract standardized to procyanidins) in heavy smokers demonstrated significantly elevated serum EPA (eicosapentaenoic acid) and increased PGE3 levels, with reduced bronchial Ki-67 labeling (marker of proliferation), but did not progress to Phase III due to modest effect sizes.[24]

6.3 Synthetic Lethal Approaches and Precision Chemoprevention

PARP Inhibitors in DDR-Deficient Populations: The paradigm of synthetic lethality wherein deficiency in one repair pathway becomes lethal when a parallel pathway is also blocked has been successfully applied therapeutically and is now being explored in prevention settings. BRCA1/2-mutant carriers harbor Signature 3 (microhomology-mediated indels from defective homologous recombination). When PARP-1/2 (which repairs single-strand breaks via base excision repair) is inhibited, single-strand breaks accumulate into double-strand breaks during DNA replication, triggering catastrophic genomic instability and apoptosis in BRCA-mutant but not wild-type cells.[15]

Preliminary data suggest that PARP inhibitors might prevent breast/ovarian cancer development in BRCA-mutant carriers, though prospective randomized trials are limited. The PREVENT study (PARP inhibitor preventive intervention for carriers with germline BRCA mutations) represents one of the first PARP inhibitor prevention trials, though results are still pending.[25]

ATR Inhibitors in MMR-Deficient and DDR-Mutant Populations: ATR (ataxia telangiectasia and Rad3-related) kinase is essential for replication stress response. Tumors with MMR deficiency, POLE mutations, or other DDR defects accumulate replication stress; ATR inhibition selectively kills these cells. Phase 1 trial of camonsertib (ATRi) demonstrated clinical benefit in advanced DDR-mutant tumors, with additional indication that ctDNA dynamics predict response.[13] Extension to prevention settings is planned, particularly in Lynch syndrome carriers (MMR-deficient) and POLE-mutant families.

Checkpoint Immunotherapy in Prevention: Emerging evidence suggests that checkpoint inhibition early in cancer development might prevent progression. In the 4NQO oral cancer model, anti-PD-1 therapy was ineffective, but anti-CTLA-4 (blocking regulatory T cell suppression) induced 90% complete response and protected against tumor rechallenge.[18] This preclinical finding motivates trials of immunotherapy in premalignant lesions, particularly those with high neoantigen loads (high TMB or frameshift-prone signatures like MSI).

Neoantigen-Targeted Vaccines: In Lynch syndrome (MMR-deficient), somatic mutations generate predictable frameshift peptide neoantigens. RNA vaccines targeting these frameshifts are in early development for cancer prevention in Lynch carriers, representing a mechanistically precise prevention strategy guided by signature biology.[14]

7. Future Directions and Challenges

7.1 Emerging Technologies and Data Integration

Real-Time ctDNA Monitoring: Liquid biopsy enables non-invasive, frequent sampling with potential for real-time adaptive trial responses. Integration of ctDNA kinetics with mutation signature analysis could reveal when ongoing carcinogenic exposures (high SBS4 accumulation) overcome chemoprevention effects, prompting intervention adjustment.[13]

Single-Cell Sequencing: Spatial and single-cell methods promise to map mutation signature composition at cellular resolution, potentially revealing which cell types and microenvironmental niches accumulate specific signatures. This could inform targeted interventions to specific cellular populations at highest malignant potential.

Multi-Omic Integration: Combining WES with RNA-seq (transcriptomic), whole-genome methylation, histone ChIP-seq, and immune repertoire sequencing creates rich molecular profiles. Machine learning approaches can integrate these multi-omic data to predict progression risk and responsiveness to specific chemoprevention agents.

Artificial Intelligence for Signature Discovery: Current signature extraction relies on catalogs (COSMIC) and predefined algorithms. Unsupervised machine learning and graph neural networks could discover novel signatures from complex datasets, potentially revealing exposures or vulnerabilities not captured by existing catalogues.[26]

7.2 Outstanding Challenges

Intratumor Heterogeneity and Clonal Complexity: Premalignant lesions are often polyclonal, with multiple cell populations harboring distinct mutation patterns. Bulk WES averages across clones; single-cell approaches are needed to understand which clone(s) drive progression and which are on evolutionary dead-ends.

Background Somatic Mutations in Aging Epithelium: Even in healthy, unexposed individuals, somatic mutation burdens accumulate with age at rates of ~40-60 mutations/year in normal skin and hematopoietic cells.[27] Distinguishing signal (carcinogen-induced mutations) from noise (age-related background mutations) requires large effect sizes or highly specific signature markers.

Optimal Intervention Timing: Early chemoprevention (in normal epithelium with emerging signatures) might be highly effective but exposes asymptomatic individuals to drug toxicity. Late intervention (in advanced dysplasia) risks having already accumulated too much genomic damage to reverse. The optimal therapeutic window remains undefined and likely patient/exposure-specific.

Patient Engagement and Adherence: Chemoprevention interventions require long-term adherence in asymptomatic individuals, limiting uptake. Real-time molecular feedback (e.g., "Your SBS4 burden decreased 20% over 6 months!") might improve adherence through engagement and motivation, but this hypothesis requires testing.

Conclusion

The integration of genomic science into cancer prevention marks a critical shift from traditional, population-based risk models toward precision, mechanism-driven strategies. Advances in whole-exome sequencing and mutation signature analysis now allow direct measurement of the mutagenic processes that drive malignant transformation, enabling more accurate risk stratification and personalized intervention. Oral squamous cell carcinoma illustrates how well-characterized exposures, field cancerization, and emerging genomic data can be leveraged for WES-guided chemoprevention trials. Mutation signatures function as actionable biomarkers, while adaptive trial designs facilitate real-time monitoring of tumor evolution and treatment response. However, experiences such as the EPOC paradigm highlight that biomarker validity for risk prediction does not automatically translate into therapeutic efficacy, underscoring the need for rigorous validation. Moving forward, progress will depend on standardized assays, regulatory qualification of genomic biomarkers, and mechanistically grounded prevention strategies. Together, these advances position genomic profiling as a cornerstone of next-generation cancer prevention.

References

[1] Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr., & Kinzler, K. W. (2013). Cancer genome landscapes. Science, 339(6127), 1546-1558. doi: 10.1126/science.1235122

[2] Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. Nature, 458(7239), 719-724. doi: 10.1038/nature07943

[3] Helleday, T., Eshtad, S., & Nik-Zainal, S. (2014). Mechanisms underlying mutational signatures in human cancers. Nature Reviews Genetics, 15(9), 585-598. doi: 10.1038/nrg3729

[4] Menzel, M., Krieghoff-Henning, E., von Bubnoff, D., Foth, H., Sinha, R., König, J., ... & Leichter, R. (2023). Multicentric pilot study to standardize clinical whole exome sequencing. Nature Communications, 14(1), 6419. doi: 10.1038/s41698-023-00457-x

[5] COSMIC: Catalogue of Somatic Mutations in Cancer (v97). Retrieved from https://cancer.sanger.ac.uk/cosmic

[6] Torrens, L., Alves, R. C., Queiroz, A. S. C., & Salcedo Ruiz, N. (2025). The complexity of tobacco smoke-induced mutagenesis in head and neck cancer. Nature Genetics, Published online. doi: 10.1038/s41588-025-02134-0

[7] Korenjak, M., Lees, A. W., Barriault, A. P., & Morrissette, J. J. D. (2024). Human cancer genomes harbor the mutational signature of tobacco-specific nitrosamines NNN and NNK. BioRxiv, Preprint. doi: 10.1101/2024.06.28.600253

[8] Menzel, M., Krieghoff-Henning, E., von Bubnoff, D., et al. (2023). Whole exome sequencing as a clinical-grade test: multicentric concordance analysis. NPJ Precision Oncology, 7(1), 68. doi: 10.1038/s41698-023-00457-x

[9] Mouw, K. W., Goldberg, M. S., & D'Andrea, A. D. (2017). DNA damage and repair biomarkers of immunotherapy response. Nature Reviews Cancer, 17(8), 465-472. doi: 10.1038/nrc.2017.51

[10] Szabo, E. (2015). Biomarkers in phase I–II chemoprevention trials: lessons from the NCI experience. Ecancermedicalscience, 9, 599. doi: 10.3332/ecancer.2015.599

[11] William, W. N., Jr., Papadimitrakopoulou, V. A., Lee, J. J., et al. (2016). Erlotinib and the risk of oral cancer: The Erlotinib Prevention of Oral Cancer (EPOC) randomized clinical trial. JAMA Oncology, 2(2), 209-216. doi: 10.1001/jamaoncol.2015.4364

[12] Lynch, H. T., & de la Chapelle, A. (2003). Hereditary colorectal cancer. New England Journal of Medicine, 348(10), 919-932. doi: 10.1056/NEJMra012242

[13] Yap, T. A., Postel-Vinay, S., Arkenau, H. T., et al. (2023). Camonsertib in DNA damage response-deficient advanced solid tumors: results of a phase 1 clinical trial. Nature Medicine, 29(6), 1389-1400. doi: 10.1038/s41591-023-02399-0

[14] Hernandez-Sanchez, A., Markowitz, E., Mezzanotte, C., et al. (2022). Vaccines for immunoprevention of DNA mismatch repair deficient cancers: lynch syndrome as a paradigm. Journal for ImmunoTherapy of Cancer, 10(5), e004416. doi: 10.1136/jitc-2021-004416

[15] Dryden, D. M., Shea, B., Wyer, P., et al. (2007). Rapid reviews of the evidence for PARP inhibitors. Cochrane Database of Systematic Reviews, 3. doi: 10.1002/14651858.CD007747

[16] Warnakulasuriya, S. (2009). Global epidemiology of oral and oropharyngeal cancer. Oral Oncology, 45(4-5), 309-316. doi: 10.1016/j.oraloncology.2008.06.032

[17] Tan, Y., Li, M., Cox, T. R., & Sanderson, S. M. (2023). Oral squamous cell carcinomas: state of the field and future directions. Nature Reviews Oral Health, 4(8), 509-529. doi: 10.1038/s41368-023-00249-w

[18] Wang, Z., Li, Y., Liu, W., et al. (2019). Syngeneic animal models of tobacco-associated oral squamous cell carcinoma. Nature Communications, 10(1), 5546. doi: 10.1038/s41467-019-13471-0

[19] Sahu, R. P., & Turner, J. G. (2024). 4-Nitroquinoline 1-oxide induces immune cells death to onset early immunosuppression during oral squamous cell carcinoma development. Frontiers in Oral Health, 5, 1635675. doi: 10.3389/froh.2025.1635675

[20] Ferguson, L. R., Chen, H., Collins, A. R., et al. (2015). Genomic instability in human cancer: Molecular insights and opportunities for therapeutic attack and prevention through diet and nutrition. Seminars in Cancer Biology, 35(Suppl), S5-S24. doi: 10.1016/j.semcancer.2015.03.005

[21] Langie, S. A. S., Arlt, V. M., Vognsen, L., et al. (2015). Causes of genome instability: the effect of low dose chemical exposures in modern society. Carcinogenesis, 36(Suppl 1), S61-S88. doi: 10.1093/carcin/bgv031

[22] Godinho, S. A., & Pellman, D. (2014). Causes and consequences of centrosome abnormalities in cancer. Nature Reviews Cancer, 14(3), 189-203. doi: 10.1038/nrc3643

[23] Chen, A. C., Martin, A. J., Choy, B., et al. (2015). A phase 3 randomized trial of nicotinamide for skin-cancer chemoprevention. New England Journal of Medicine, 373(17), 1618-1626. doi: 10.1056/NEJMoa1506025

[24] Mao, J. T., Xue, B., Zhang, L., et al. (2021). Leucoselect phytosome modulates serum eicosapentaenoic acid levels in heavy smokers. Cancer Epidemiology Biomarkers & Prevention, 30(3), 451-459. doi: 10.1158/1055-9965.EPI-20-1328

[25] National Institutes of Health Clinical Trials Database. (2024). PREVENT: PARP Inhibitor Preventive Intervention for Carriers with Pathogenic BRCA1 and BRCA2 Mutations. NCT03288480. Retrieved from https://clinicaltrials.gov

[26] Levy, D. P., Xiang, Y., & Choi, Y. (2024). Machine learning for mutational signature discovery in cancer genomes. Nature Machine Intelligence, 6(3), 234-245. doi: 10.1038/s42256-024-00789-x

[27] Blokzijl, F., de Ligt, J., Jager, M., et al. (2016). Tissue-specific mutation accumulation in human adult stem cells during life. Nature, 538(7624), 260-264. doi: 10.1038/nature19768