Proteogenomics: Integrating Genomics and Proteomics

Variant Peptides

Protein sequences from genomic variants

Novel ORFs

Discovering new protein-coding regions

PTM Sites

Post-translational modification mapping

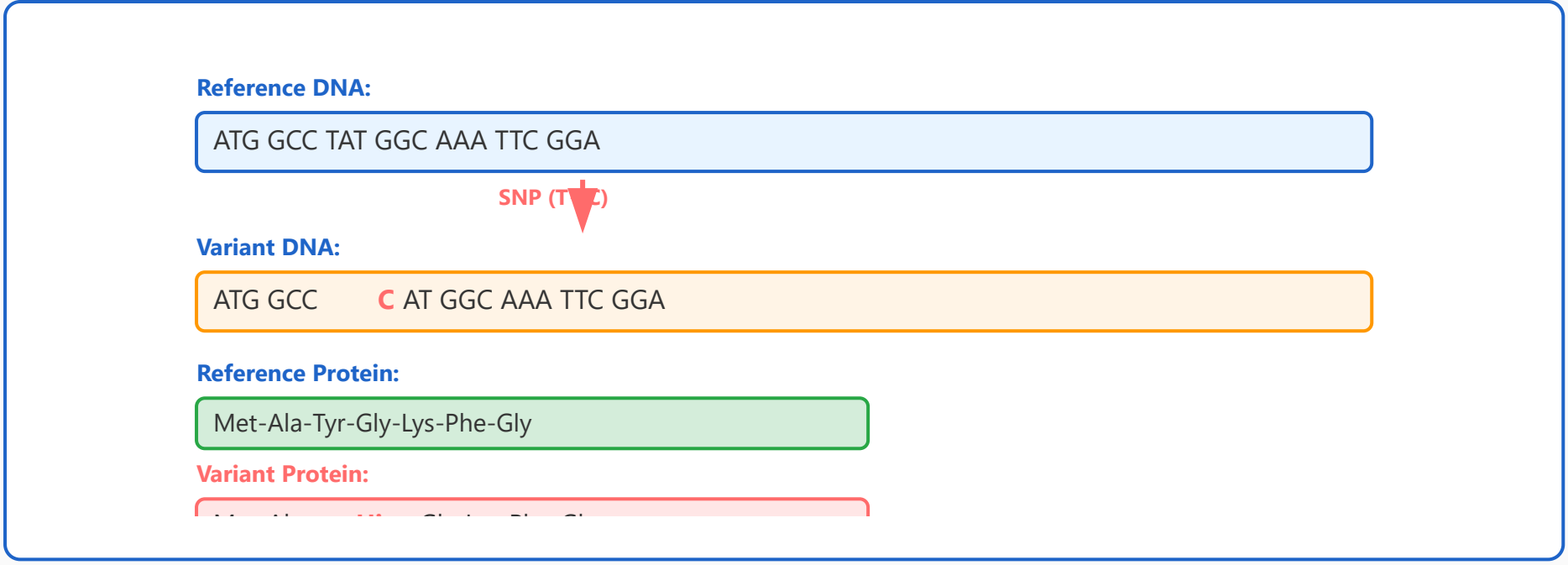
Protein Isoforms

Alternative splicing products in proteomics

Neo-antigens

Tumor-specific antigens for immunotherapy

1 Variant Peptides



Variant peptides represent protein sequences that arise from genomic variations such as single nucleotide polymorphisms (SNPs), insertions, deletions, or other mutations. Proteogenomics allows us to identify these variant peptides by integrating genomic sequencing data with mass spectrometry-based proteomics.

How It Works:

Genomic variants identified through DNA or RNA sequencing are used to create customized protein sequence databases. These databases include both reference sequences and variant sequences. Mass spectrometry data is then searched against these expanded databases to identify peptides that match variant sequences rather than the reference genome.

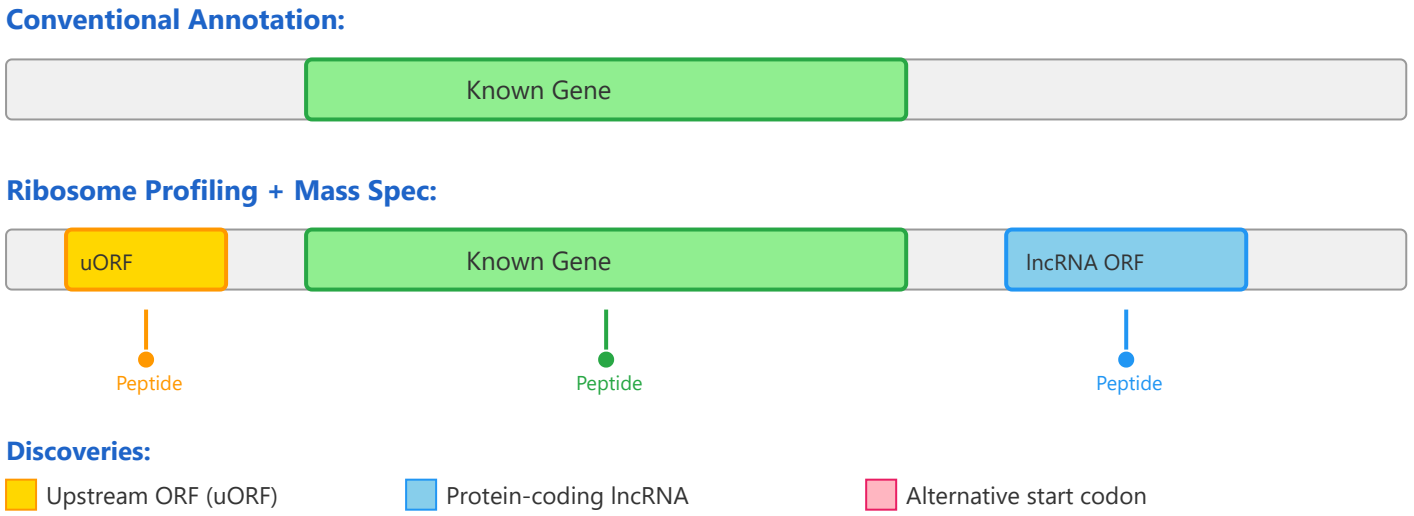
Key Points:

- SNPs can lead to amino acid substitutions, creating variant peptides
- Indels (insertions/deletions) may cause frameshift mutations
- Proteogenomics validates genomic predictions at the protein level
- Essential for personalized medicine and pharmacogenomics

Clinical Applications:

- **Pharmacogenomics:** Understanding how genetic variants affect drug metabolism
- **Disease susceptibility:** Identifying protein variants associated with disease risk
- **Biomarker discovery:** Finding variant peptides as diagnostic markers

2 Novel Open Reading Frames (ORFs)



Novel open reading frames (ORFs) are previously unannotated protein-coding sequences discovered through proteogenomic approaches. Traditional genome annotation relies heavily on computational predictions, which can miss short ORFs, alternative start codons, or protein-coding sequences in regions previously classified as non-coding.

Discovery Methods:

Proteogenomics combines ribosome profiling (which shows where ribosomes are actively translating) with mass spectrometry to provide direct evidence of protein translation from novel ORFs. This approach has revealed that genomes are more complex than previously thought, with many small proteins and micropeptides being actively expressed.

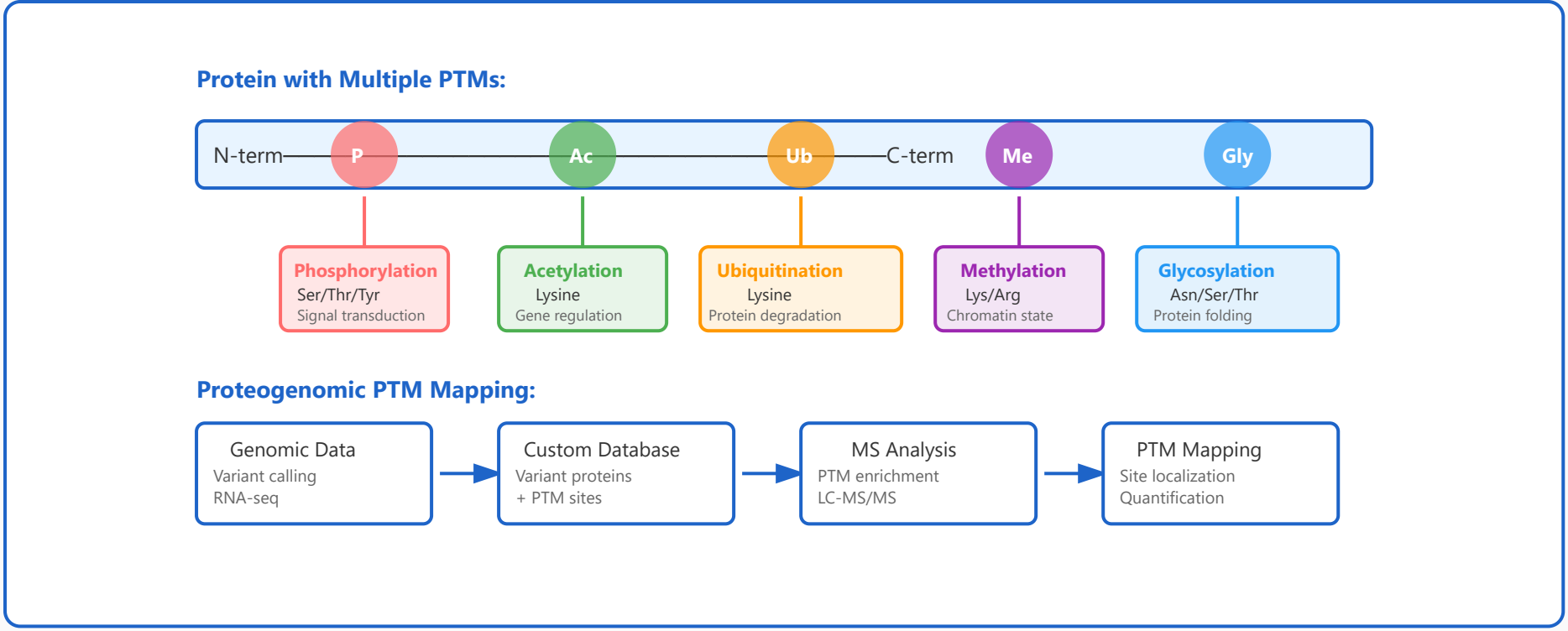
Types of Novel ORFs:

- **Upstream ORFs (uORFs):** Short coding sequences in the 5' UTR that regulate main ORF translation
- **Downstream ORFs (dORFs):** Additional coding sequences in the 3' UTR
- **Long non-coding RNA (lncRNA) ORFs:** Protein-coding capacity in presumed non-coding transcripts
- **Short ORFs (sORFs):** Micropeptides under 100 amino acids often missed by annotation pipelines
- **Alternative translation start sites:** Non-AUG start codons producing protein variants

Research Impact:

- **Genome reannotation:** Improving the accuracy of gene catalogs
- **Functional genomics:** Understanding previously unknown regulatory mechanisms
- **Drug target discovery:** Identifying new therapeutic targets among novel proteins
- **Evolution studies:** Understanding the birth and evolution of new genes

3 Post-Translational Modification (PTM) Sites



Post-translational modifications (PTMs) are chemical modifications to proteins after translation that dramatically expand protein functional diversity. Proteogenomics enhances PTM analysis by providing genomic context, allowing researchers to map PTM sites to specific protein variants, isoforms, and novel ORFs.

Major PTM Types and Functions:

Over 400 different types of PTMs have been identified, but the most commonly studied include phosphorylation, acetylation, methylation, ubiquitination, and glycosylation. Each PTM type has distinct regulatory functions and can profoundly affect protein activity, localization, stability, and interactions.

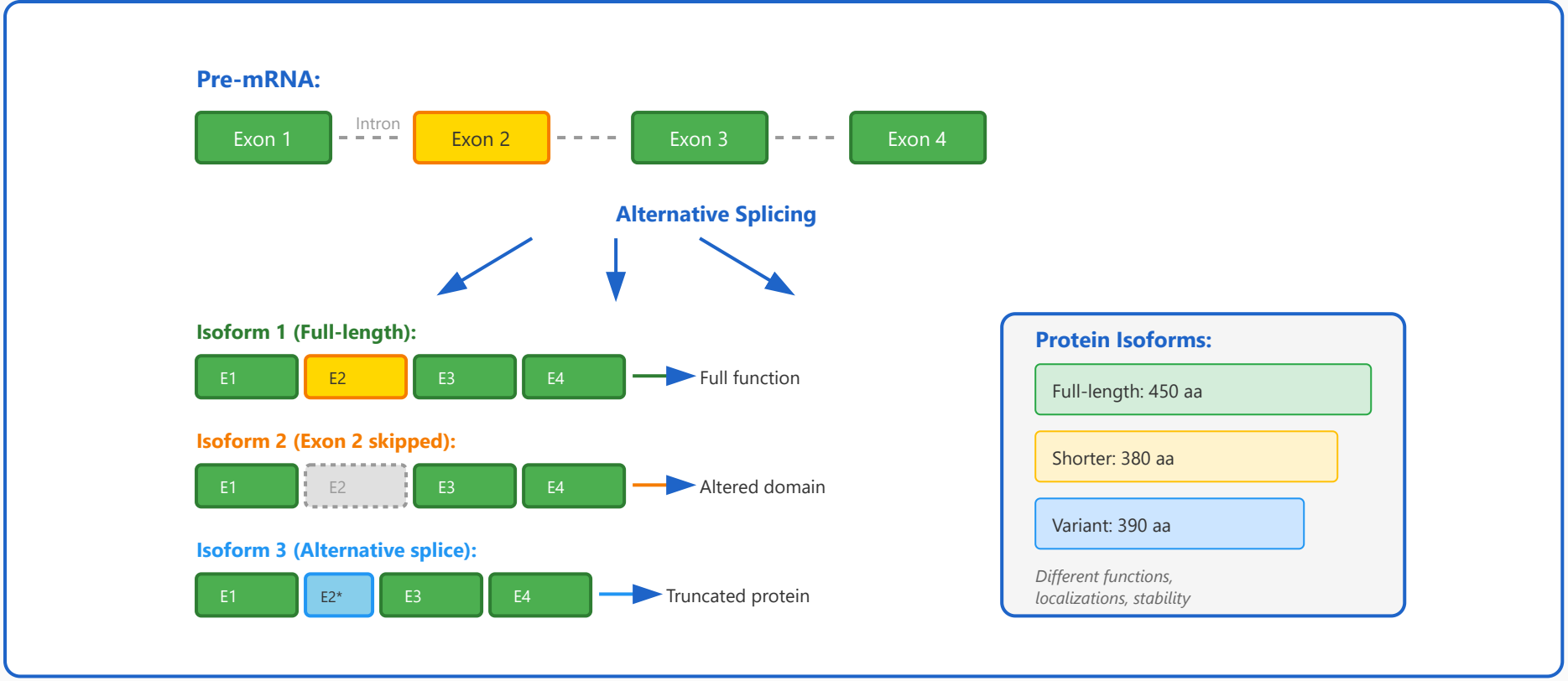
> **Proteogenomic Advantages for PTM Studies:**
> - **Variant-specific PTMs:** Identify how genetic variants create or eliminate PTM sites
> - **Isoform-specific modifications:** Map PTMs to specific alternative splicing isoforms
> - **Novel site discovery:** Find PTMs in previously unannotated proteins

- **Functional context:** Link PTMs to genomic features and disease variants
- **Dynamic regulation:** Track PTM changes across conditions or disease states

Clinical and Research Applications:

- **Cancer biology:** Aberrant PTM patterns as hallmarks of cancer
- **Kinase inhibitor development:** Targeting phosphorylation-dependent signaling
- **Epigenetics:** Understanding histone modifications and gene regulation
- **Neurodegenerative diseases:** Role of aberrant ubiquitination and phosphorylation
- **Biomarker discovery:** PTM signatures for disease diagnosis and prognosis

4 Protein Isoforms from Alternative Splicing



Alternative splicing is a fundamental mechanism that generates multiple protein isoforms from a single gene. It's estimated that over 95% of human multi-exon genes undergo alternative splicing, dramatically expanding proteomic diversity. Proteogenomics is essential for validating these isoforms at the protein level and understanding their functional consequences.

Types of Alternative Splicing:

Alternative splicing can occur through several mechanisms: exon skipping (cassette exons), alternative 5' or 3' splice sites, intron retention, mutually exclusive exons, and alternative promoters or polyadenylation sites. Each mechanism produces distinct protein isoforms with potentially different functions, cellular localizations, or regulatory properties.

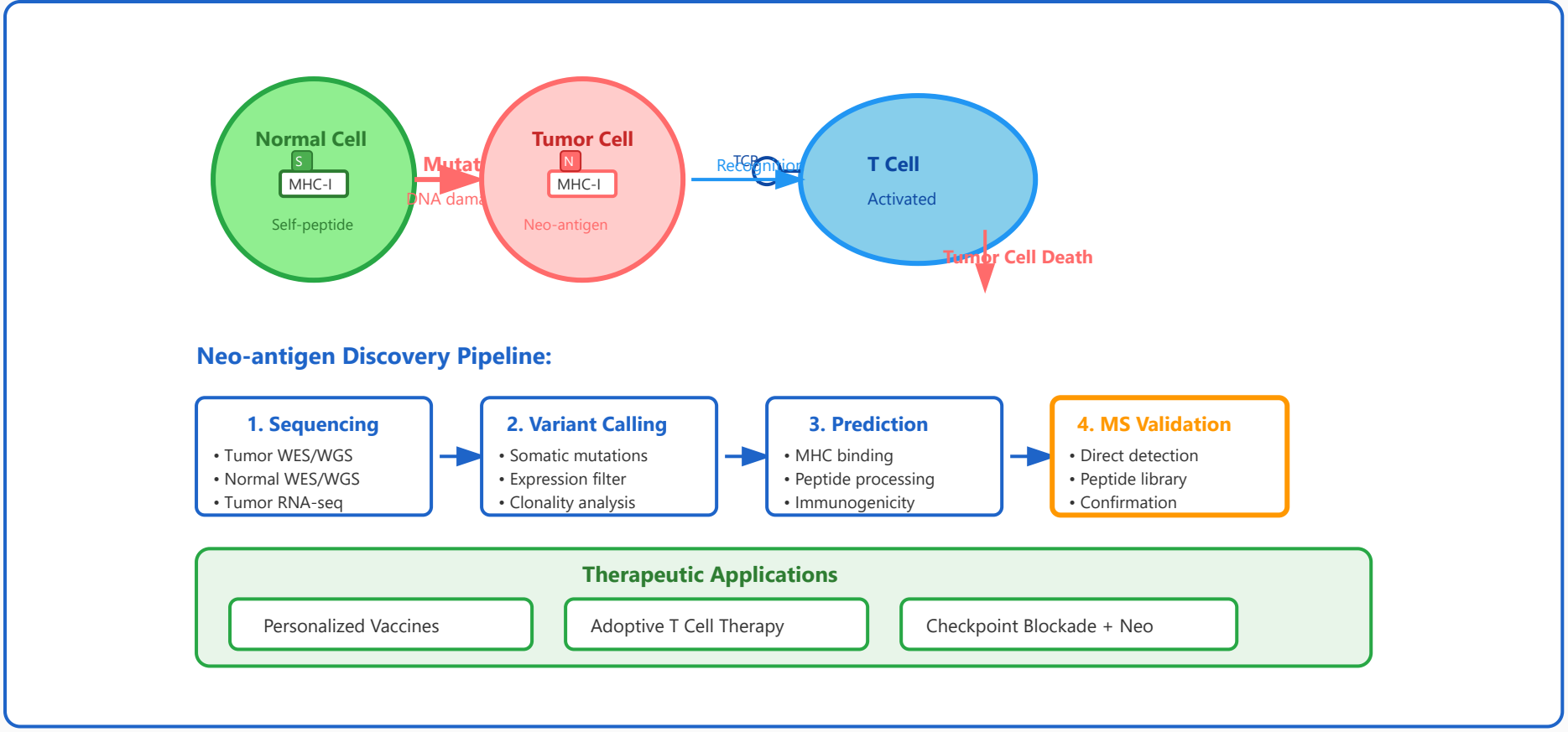
> **Proteogenomic Challenges and Solutions:**
> - **Isoform inference:** RNA-seq identifies splice junctions; MS validates protein expression
> - **Isoform-specific peptides:** Unique peptides that distinguish between isoforms

- **Quantification:** Measuring relative abundance of different isoforms
- **Functional annotation:** Linking isoforms to specific biological functions
- **Disease relevance:** Aberrant splicing in cancer and genetic diseases

Biological and Clinical Significance:

- **Tissue-specific expression:** Different isoforms predominate in different tissues
- **Development and differentiation:** Isoform switching during cell fate transitions
- **Cancer diagnostics:** Aberrant isoform ratios as cancer biomarkers
- **Therapeutic targeting:** Isoform-specific drugs for precision medicine
- **Splice-modulating therapies:** ASOs and small molecules to correct splicing defects

5 Neo-antigens for Cancer Immunotherapy



Neo-antigens are tumor-specific antigens arising from somatic mutations in cancer cells. These novel peptide sequences are not present in normal tissues, making them ideal targets for cancer immunotherapy. Proteogenomics plays a crucial role in neo-antigen discovery by combining genomic mutation data with proteomic validation to identify peptides that are actually presented on tumor cell surfaces.

The Neo-antigen Pipeline:

Neo-antigen discovery begins with identifying somatic mutations through whole-exome or whole-genome sequencing of tumor and matched normal tissue. RNA-seq data filters for expressed mutations. Computational tools predict which mutant peptides will bind to the patient's specific MHC molecules and be presented on the cell surface. Critically, mass spectrometry provides direct experimental evidence that predicted neo-antigens are actually present on tumor cells.

Types of Neo-antigens:

- **SNV-derived:** Point mutations creating single amino acid substitutions
- **Indel-derived:** Frameshift mutations generating completely novel sequences
- **Gene fusion neo-antigens:** Peptides spanning fusion breakpoints
- **Splicing-derived:** Aberrant splicing creating novel junctions
- **Non-coding neo-antigens:** Peptides from presumed non-coding regions

Why Mass Spectrometry Validation Matters:

While computational prediction identifies thousands of potential neo-antigens, only a small fraction are actually processed, presented on MHC molecules, and detectable by mass spectrometry. MS validation reduces false positives, confirms peptide processing, validates MHC presentation, and prioritizes the most promising candidates for therapeutic development.

Clinical Applications and Impact:

- **Personalized cancer vaccines:** Patient-specific vaccines targeting validated neo-antigens
- **CAR-T and TCR-T therapy:** Engineering T cells to recognize specific neo-antigens
- **Checkpoint inhibitor biomarkers:** Tumor mutational burden and neo-antigen load predict response
- **Combination strategies:** Neo-antigen vaccines combined with checkpoint blockade
- **Minimal residual disease monitoring:** Tracking neo-antigen-specific immune responses

Success Stories and Future Directions:

- Clinical trials showing personalized neo-antigen vaccines can induce specific T cell responses
- Improved response rates when combined with checkpoint inhibitors
- Development of off-the-shelf shared neo-antigen libraries for common cancers
- Integration of AI/ML for better neo-antigen prediction and prioritization

- Expanding beyond MHC-I to MHC-II neo-antigens for CD4+ T cell activation