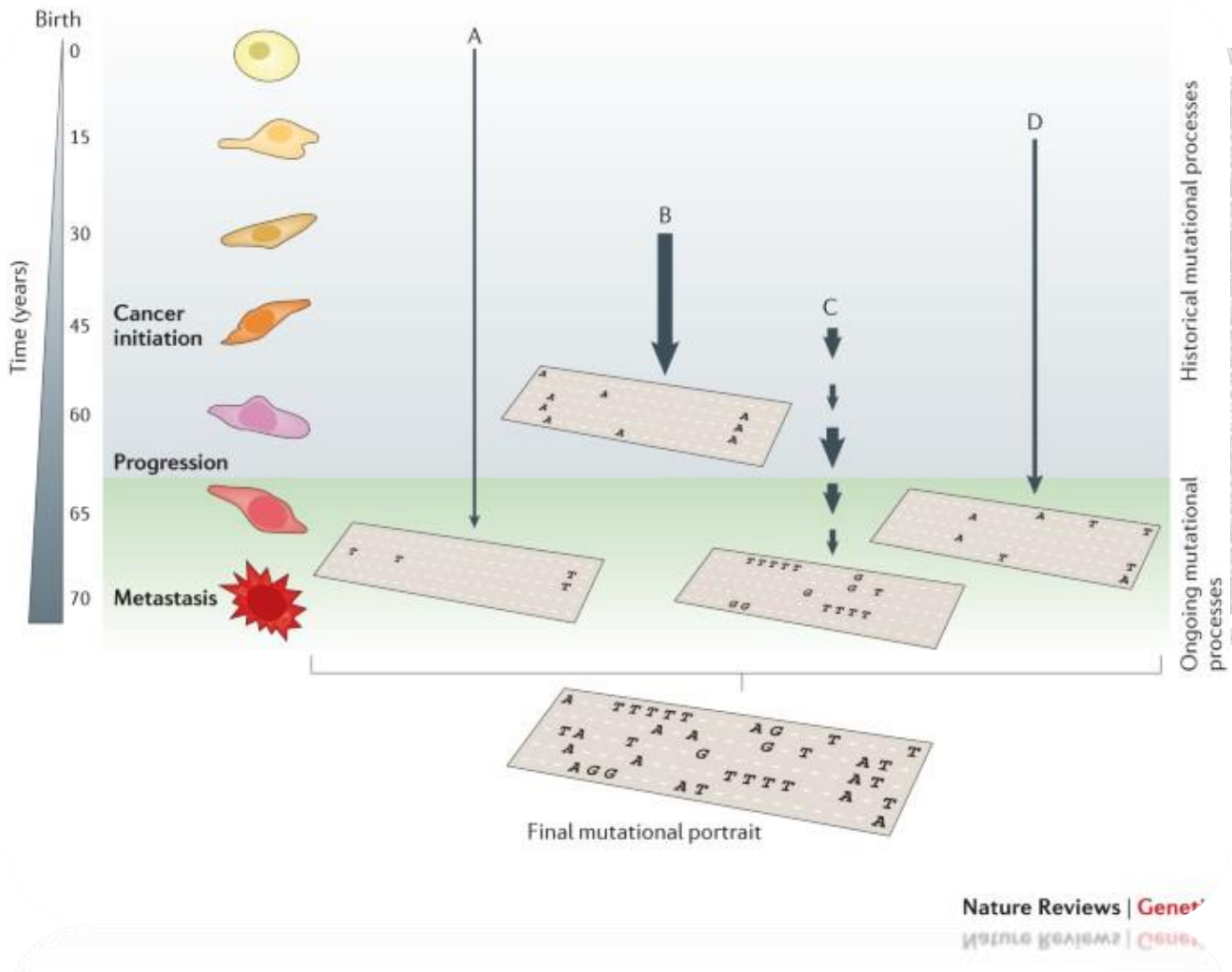



# MUTATIONAL SIGNATURES

A M Mahedi Hasan, PhD | CBiol MRSB | FHEA
Senior Research Associate (Computational Biology
Treatment Resistance Group
UCL Cancer Institute

# Mutational signature

- a physiological readout of the biological history of a cancer

- Mutational signatures provides an account of the mechanism and the degree to which it has been affected by this perturbation (historic and ongoing).
- Driver mutations

  Somatic mutations that are causally implicated in oncogenesis and that confer selective advantages during the evolution of a cancer.

- Passenger mutations

  may not be causative of cancer development but that are a rich source of historical information.

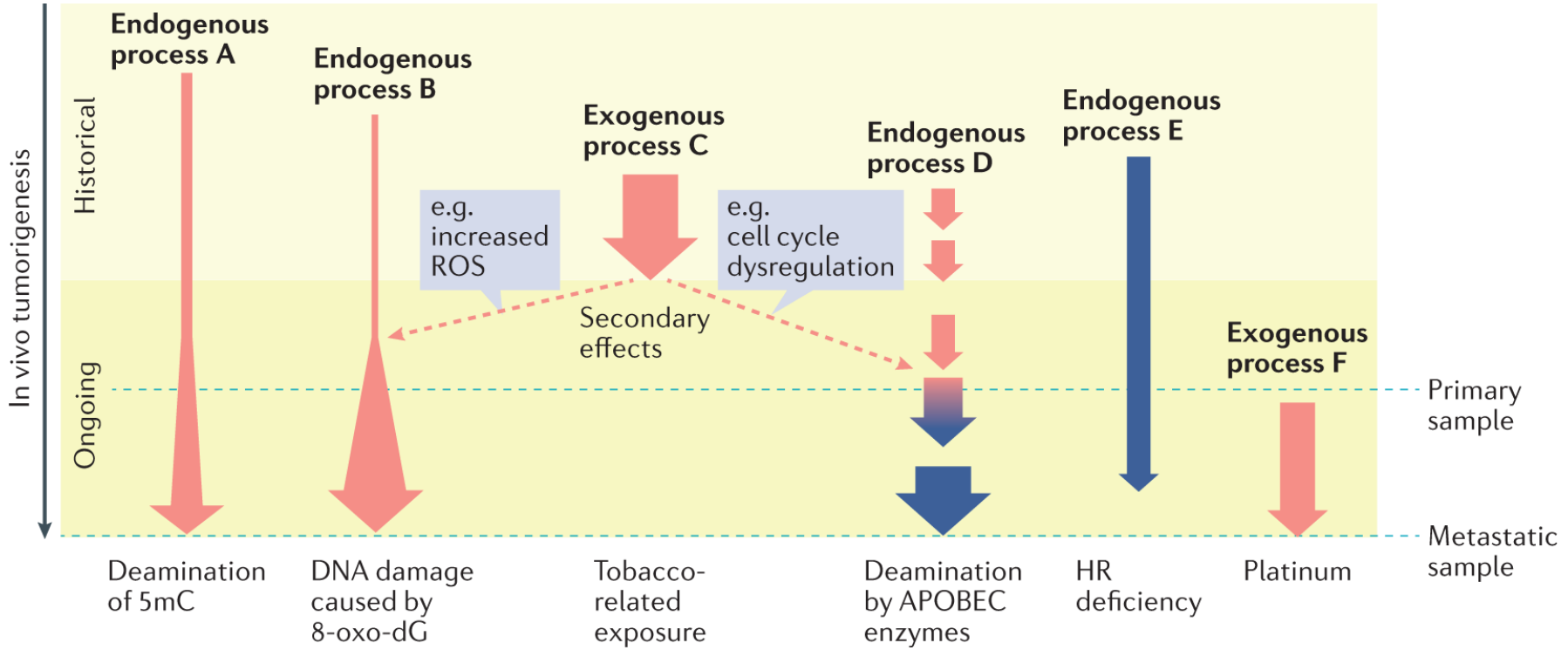
# Mutational signature

- a physiological readout of the biological history of a cancer

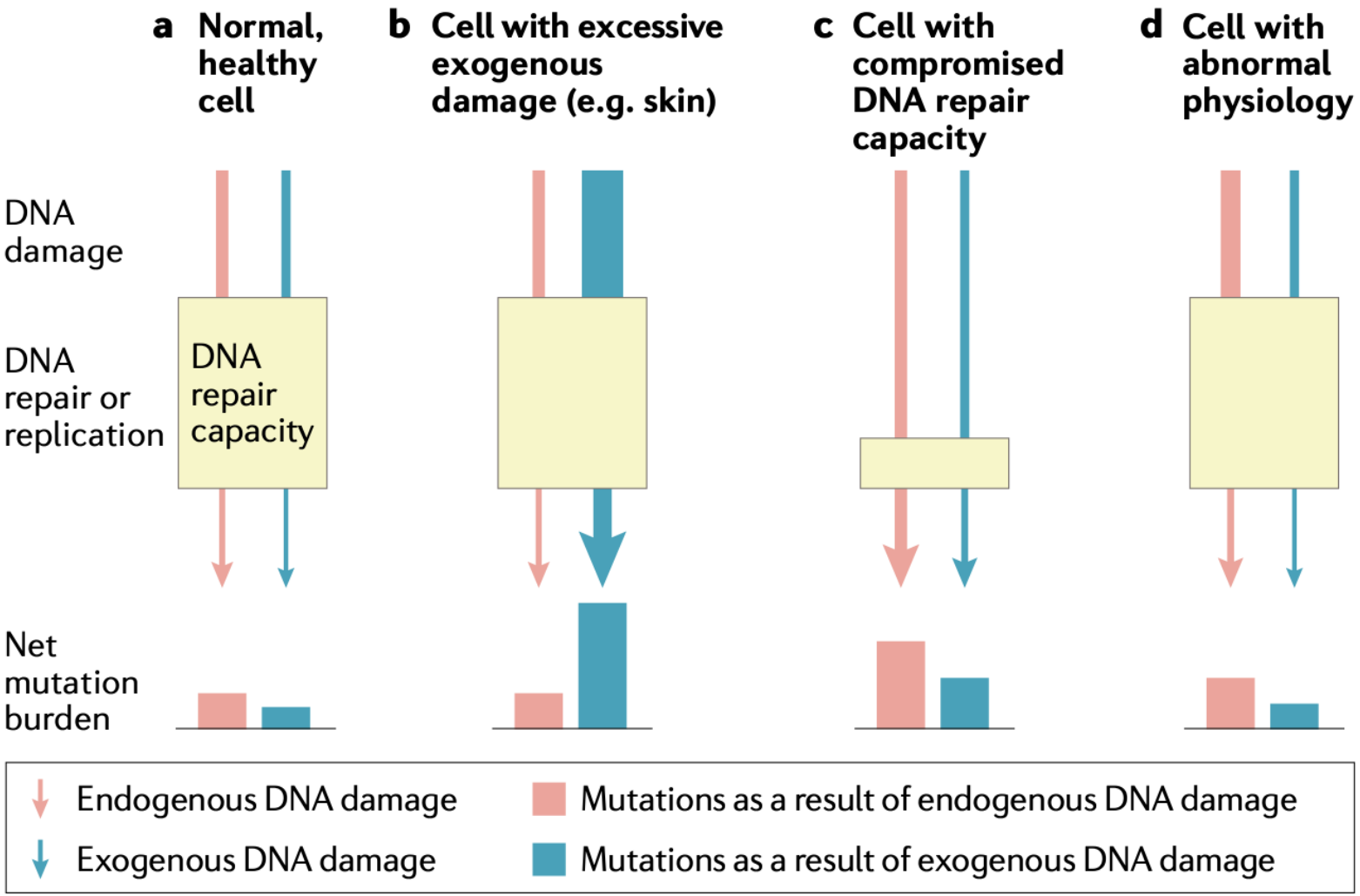
"The collective somatic mutations observed in a cancer are the outcome of multiple mutagenic processes that have been operative over the lifetime of a patient."

- Each mutagenic process leaves behind –
  - A mutational signature, which comprises –
    - Type of DNA damage(s)
    - DNA repair process(s)
  - In relation to exogenous and endogenous DNA damaging agents, as well as by the DNA replicative mechanisms
  - The result is –
    - Base substitution
    - Insertions and deletions
    - Structural variation

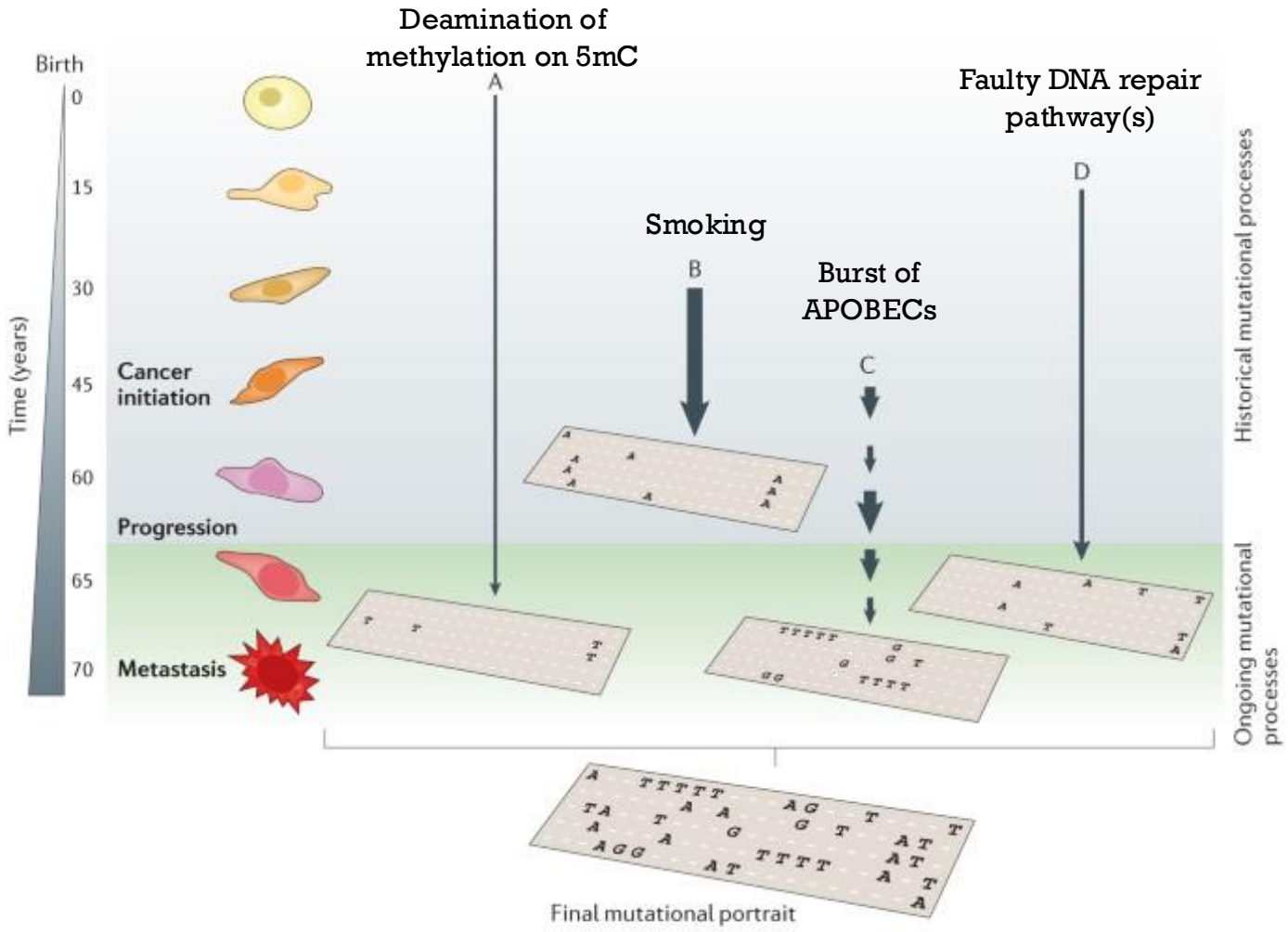
# Mutational processes can be endogenous or exogenous



# The extent of mutational process leaves its signature behind



# The final mutational portrait is the sum of all of the different mutational processes

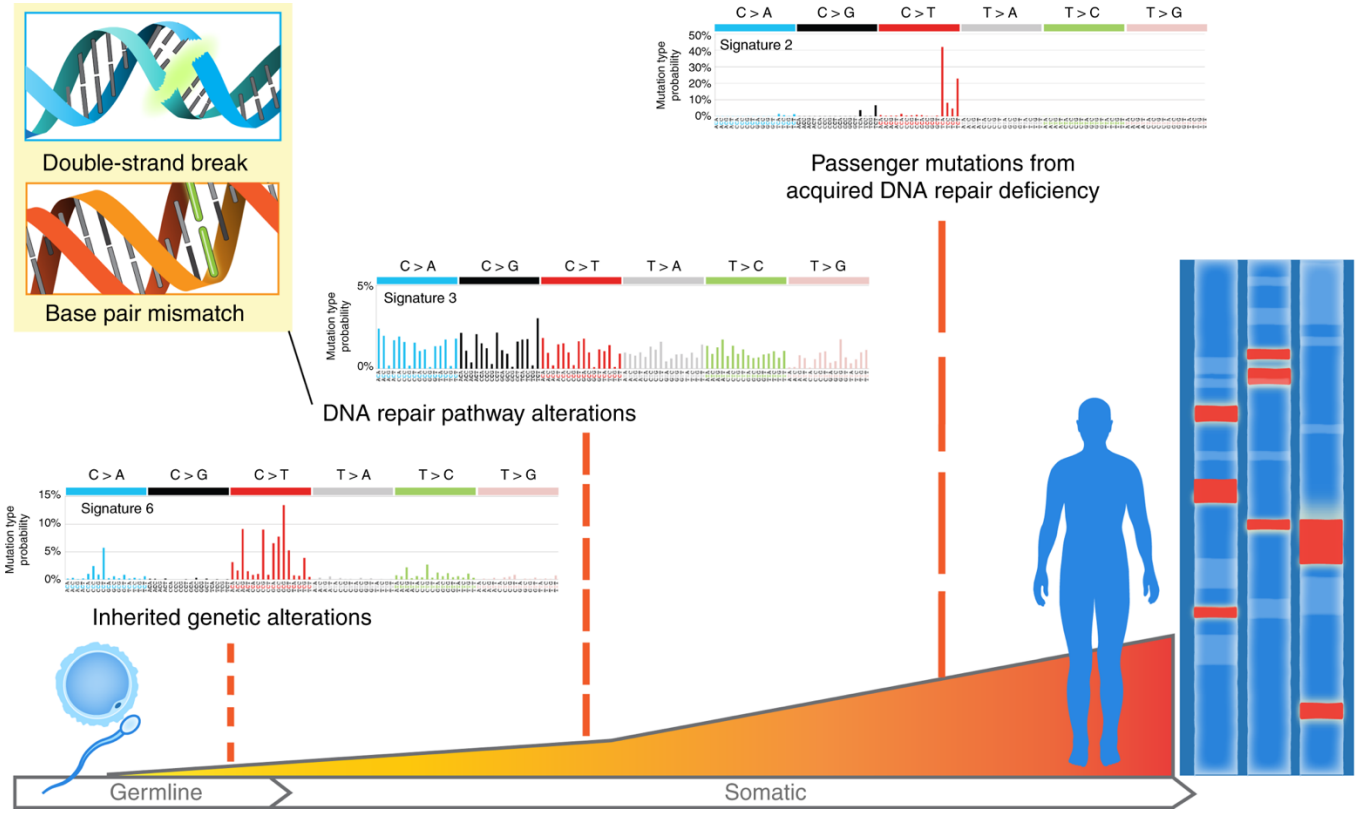



Likely clonal
Pre-metastatic

Clonal and/or subclonal

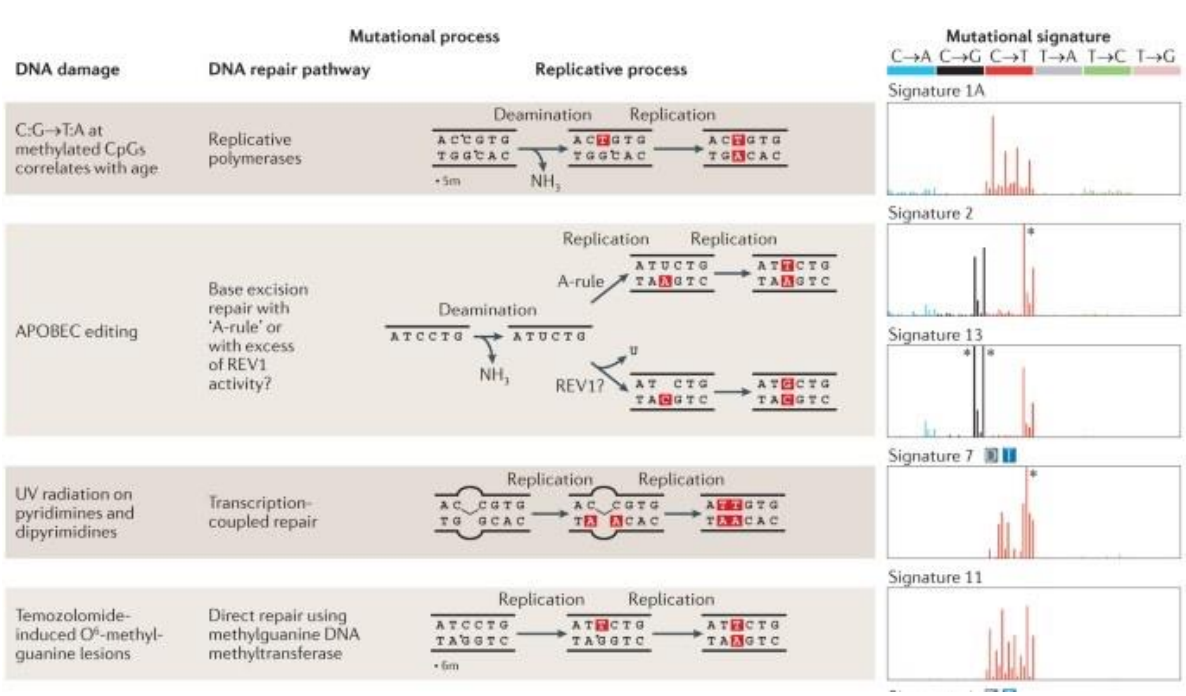
- Historical mutational processes are no longer active.
- Ongoing mutational processes reflect active biological processes in the cancer that could be exploited either as biomarkers to monitor treatment response or as therapeutic anticancer targets.

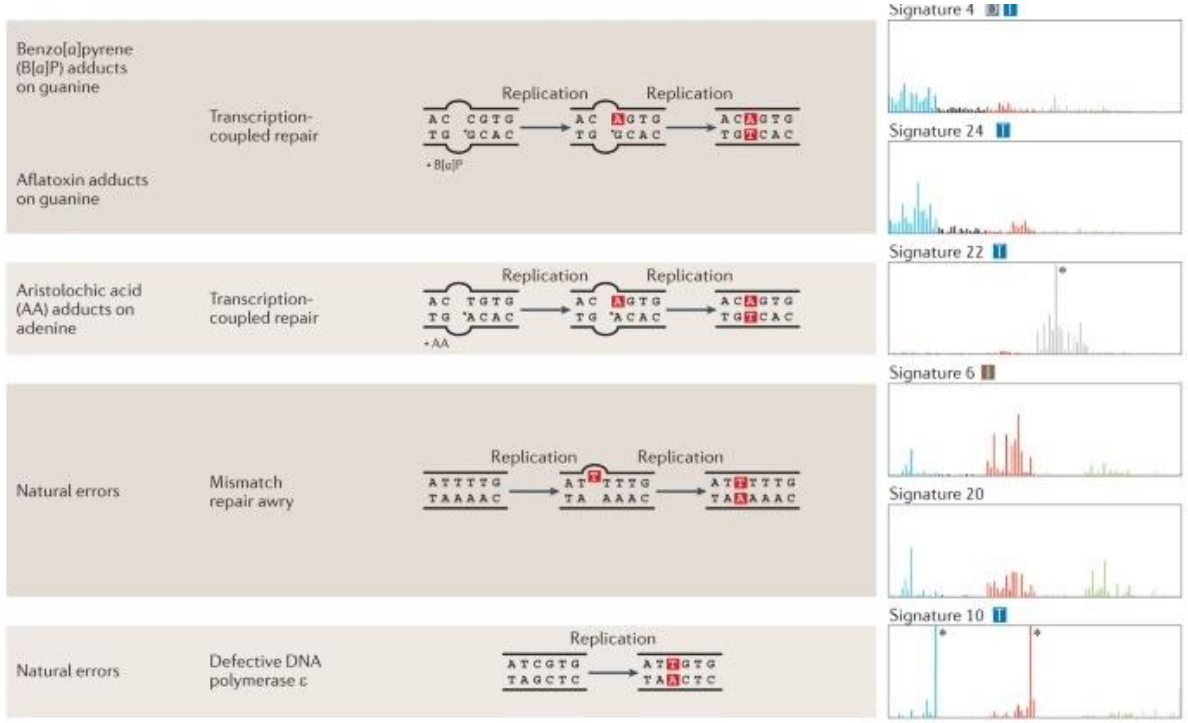
# Mutational processes can be inherited or acquired during oncogenesis



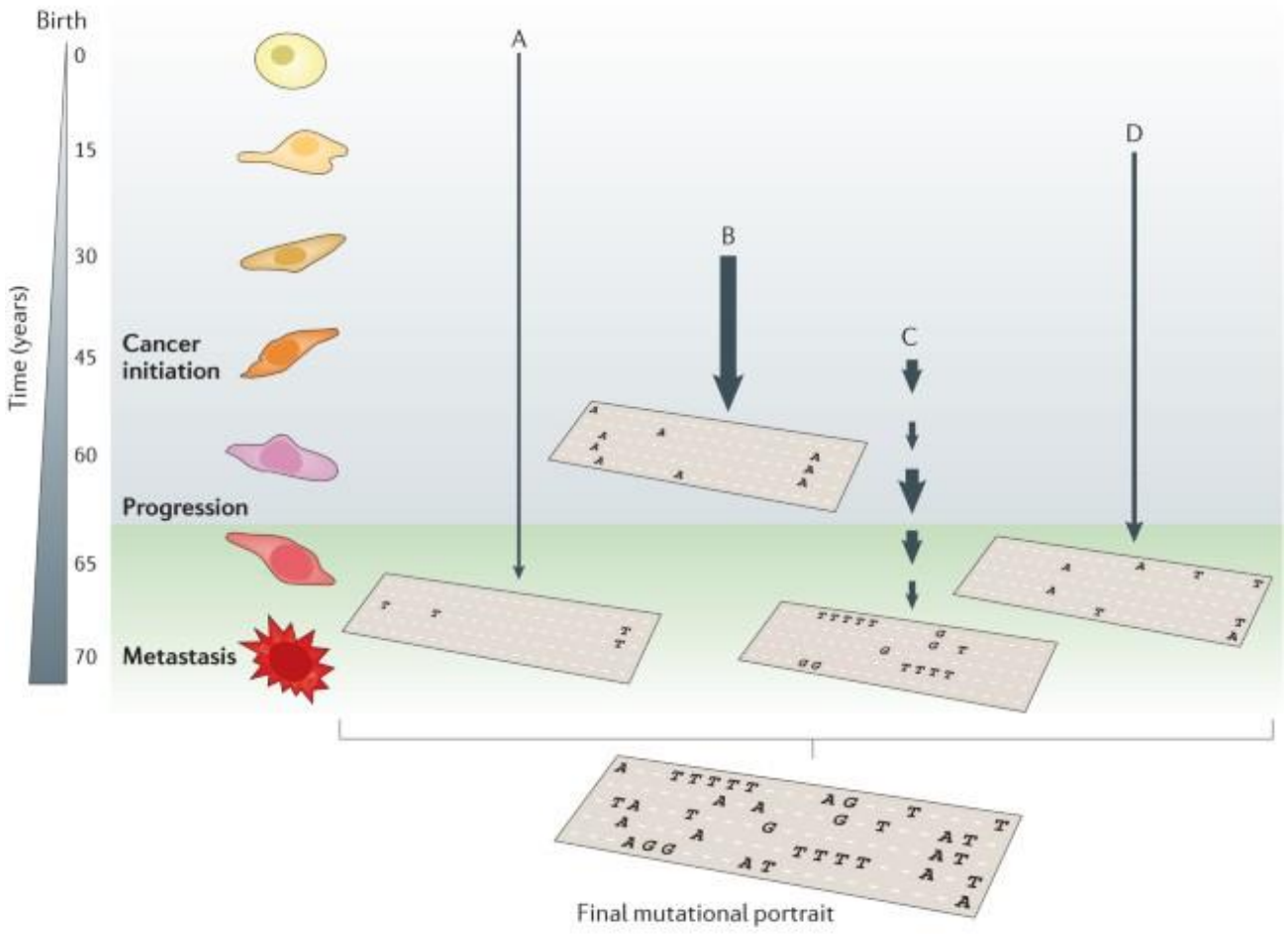

- An individual's unique mutational signature is a record of the types of DNA alterations sustained throughout their lifetime
- Can be studied to identify unique patterns of etiology-specific alterations, including carcinogens or DNA repair pathway defects, the latter of which can be inherited or acquired during oncogenesis.

# A few mutational signatures and their corresponding mutations processes



# The algorithm for determining mutational signature




**C**atalogue ≈ **S**ignature × **E**xposure

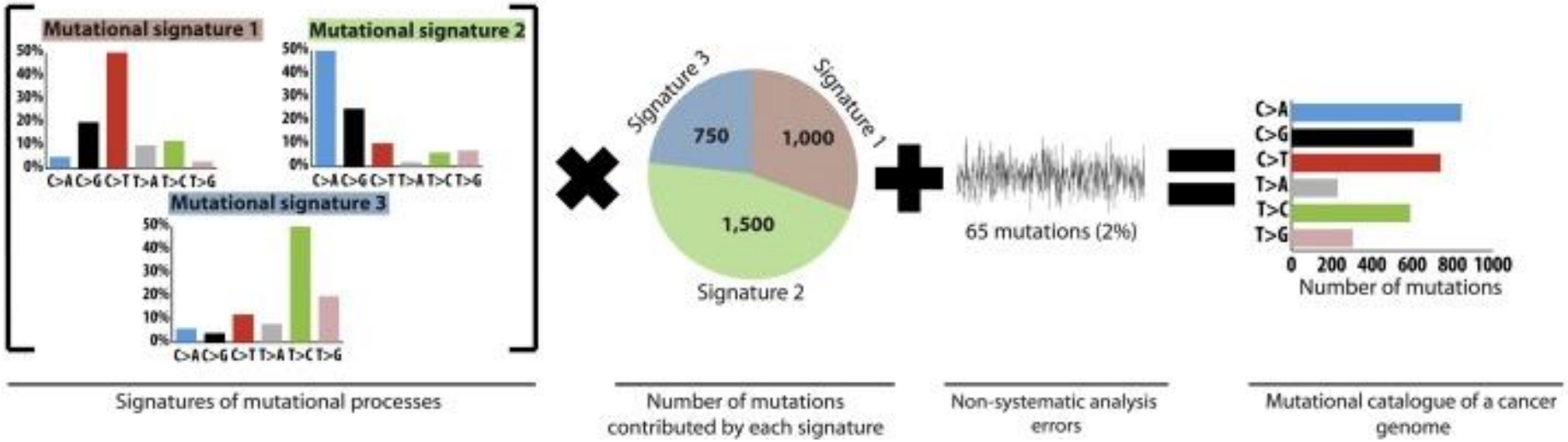
**E**xposure (total number of mutations is a proxy of the amount of each signature)

**S**ignature

**C**atalogue (final mutational profile)

# The process of determining Mutational Signature

## SIGNATURE × EXPOSURE ≈ CATALOGUE



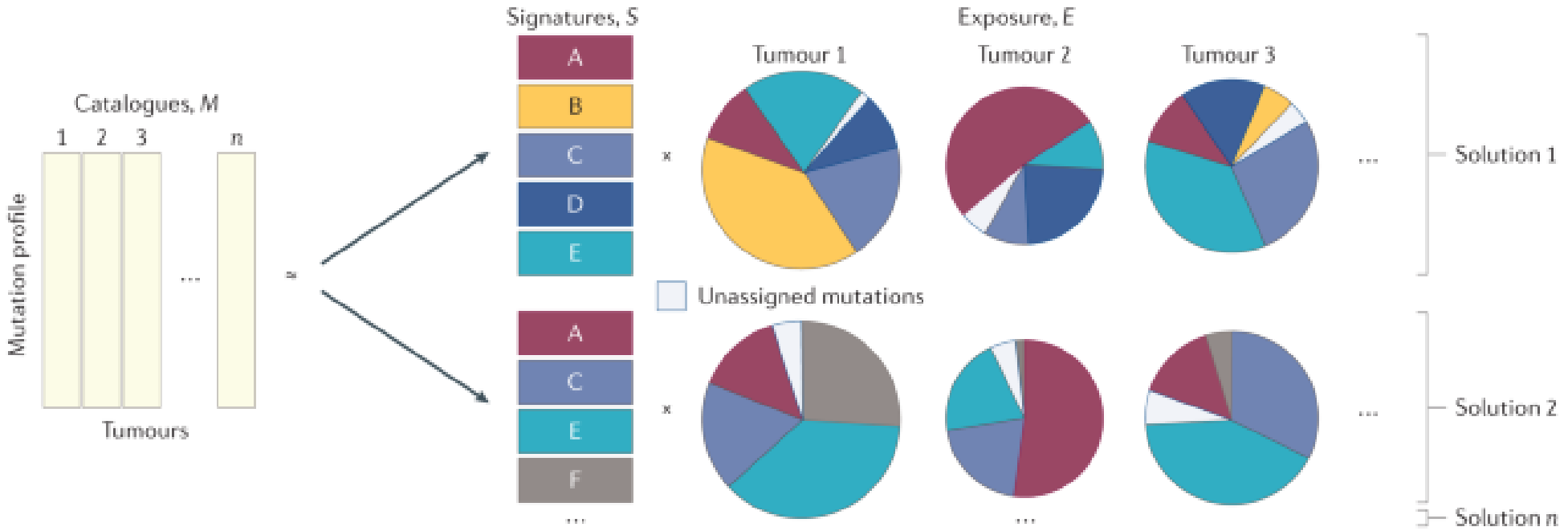

### Extraction

Estimating the number of signatures and their profiles

### Assignment

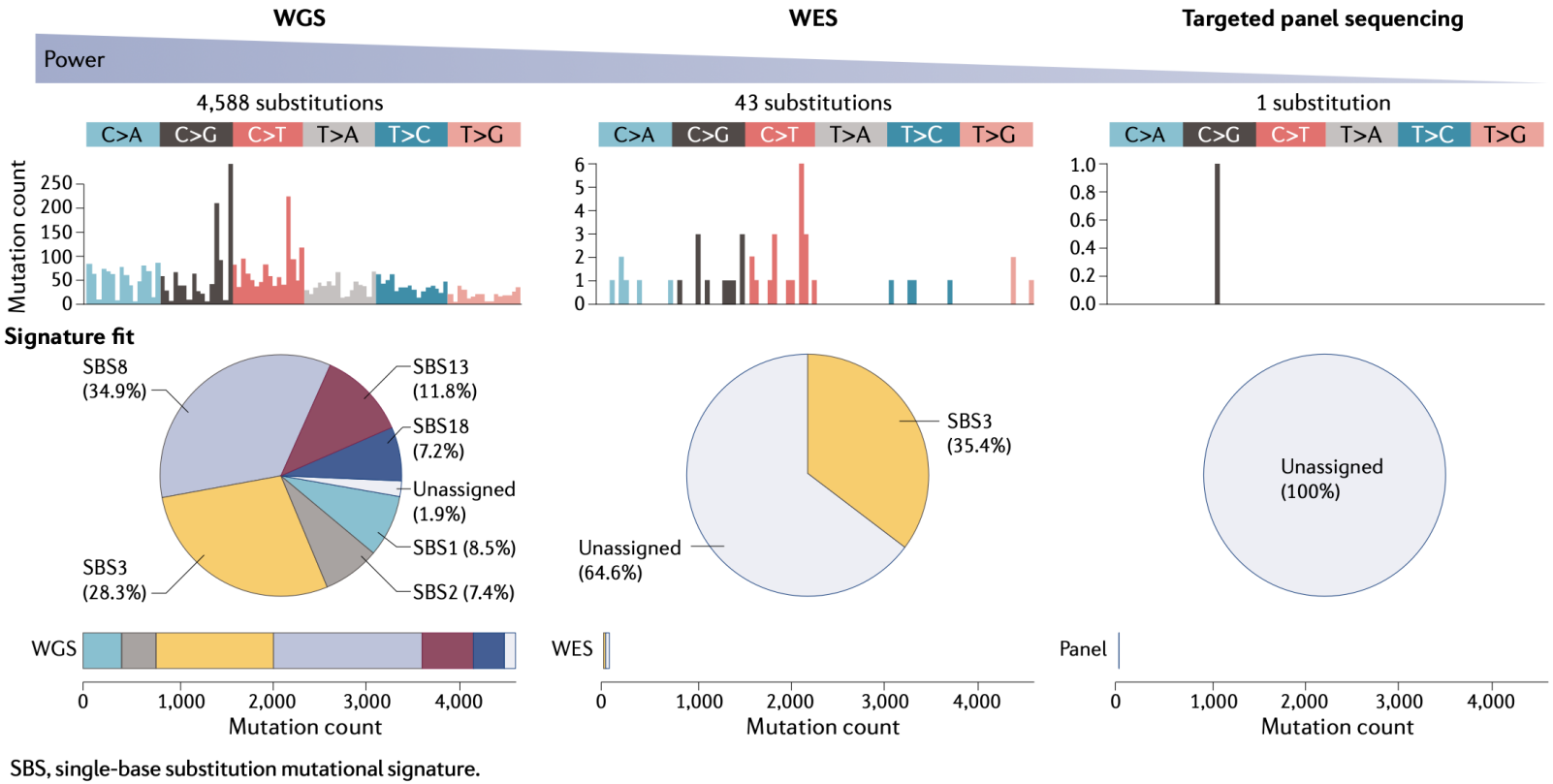
Estimating how much of each of the signatures is present in each sample

# Determining mutational signatures is a multiple-solution problem




$$M \approx S \times E$$

# Power for signature detection with different sequencing approaches




SBS, single-base substitution mutational signature.

- In general, whole cancer genomes have thousands of mutations, hundreds of insertions or deletions (indels), and tens to hundreds of rearrangements. by contrast, exome-sequenced cancers have only ~1–3% of the human genome footprint (depending on the bait set used), and therefore orders of magnitude fewer mutations.
- A breast cancer genome (PD6413a), whole-genome sequencing (WGS) at 40-fold depth revealed 4,588 substitutions.
- A whole-exome sequencing (WES) experiment with the same sample revealed 43 mutations using 96 channels. When 43 mutations are distributed across 96 channels, the numbers of mutations may be so low as to result in many channels with counts of 0 or 1.
- Targeted sequencing has an even smaller genomic footprint, and here a typical truSight oncology 500 (Illumina) targeted panel would have detected one substitution.
- For WES and targeted sequencing data, indels may be down to single digits, and rearrangements would not be reported at all.

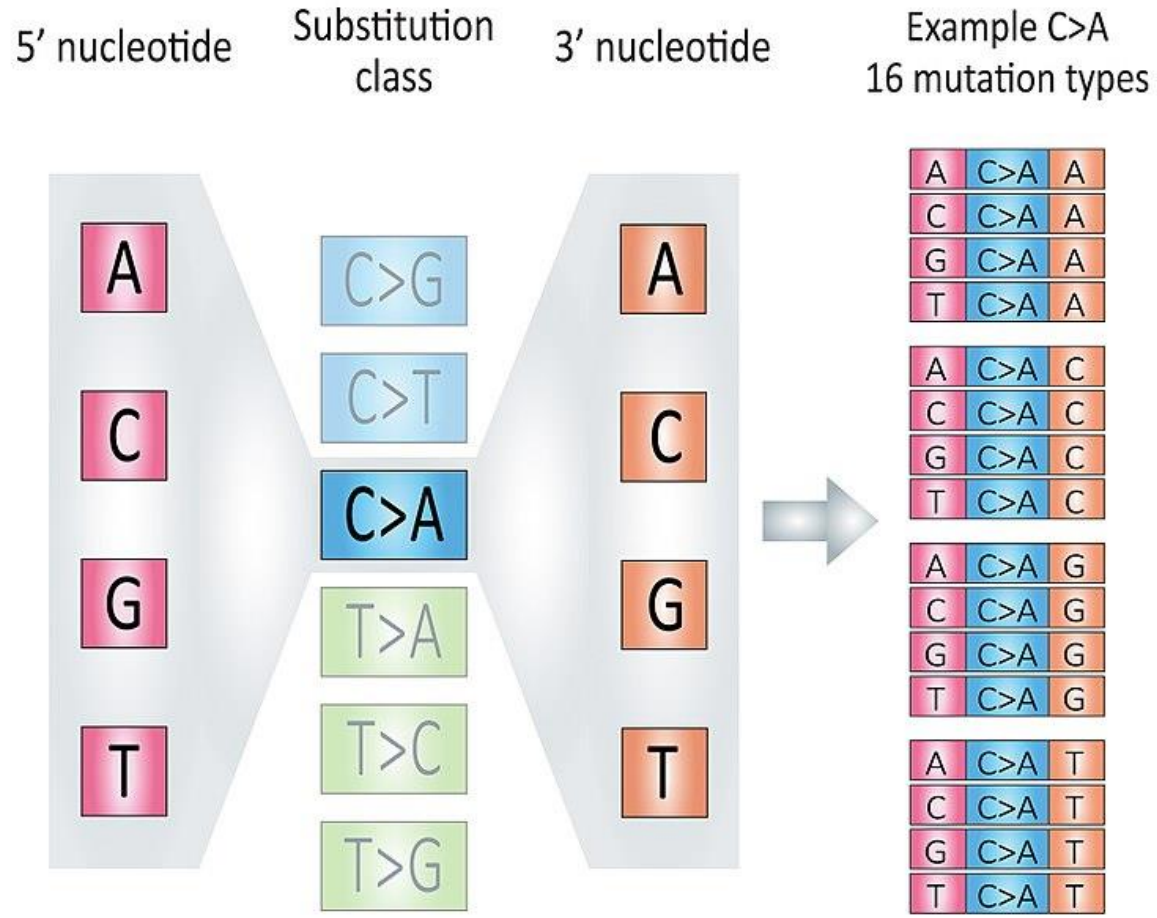
# Approaches

- Global approach of mutational signature -
  - Irrespective of the tissue type, signatures from all cancer types are aggregates and averaged to derive a set of consensus signatures
  - Presuming that more samples provide more power for discerning new signatures
  - Ignoring possible tissue-specific signature properties that reflect organ-specific biology (which could be probable)

- Local approach of mutational signature
  - Restricting signature extraction within individual tissue types.
  - As an imbalanced tissue sampling resulted in certain tissue types being more influential and thereby introduces potential bias.
  - Possible to locally extracted signatures between different organs.

# Single Base Substitution signature

## 6 to 96 channels




+ 80 more

# Single Base Substitution signatures

- Regardless of the algorithms used for signature identification, common signatures tend to be consistently identifiable in most cohorts examined -
  - SBS1, caused by deamination of 5-methylcytosine.
- Signatures associated with environmental exposures tend to be immediately demonstrable -
  - UV-associated SBS7 and DBS1.
- Deficiencies in specific DNA repair pathways produce marked mutagenesis.
  - mismatch repair (MMR) deficiency associated SBS26 and SBS44
- Some endogenous signatures are highly distinctive and easily discernible,
  - Apolipoprotein B mRNA editing enzyme, catalytic polypeptide like (APOBEC)-associated SBS2 and SBS13.
- Rare mutational processes that are present at low population frequencies may be more challenging to extract and will reveal themselves only if they are present in the specific cohort examined
  - treatment associated (iatrogenic) signatures.
  - Several othher signatures were noted as treatment related. For example, SBS31 and SBS35, associated with platinum; SBS90 attributed to duocarmycin) in COSMIC v3 collection.

# ID signatures

- Indels are common in cancers, occurring at ~10% of the frequency at which substitutions occur, and their genomic locations and sequence compositions are non-random.
- Indels cannot always be pinned to a defined coordinate with the same precision as substitutions because it is impossible to pinpoint the deleted or inserted position in a polynucleotide repeat tract. Thus, indels have been classified more simply, on the basis of their -
  - type (deletion, insertion or complex),
  - size and
  - whether there are features at indel junctions that could reveal biological underpinnings
- 1bp indels occurring at repetitive tracts commonly arise from strand slippage during replication, whereas indels that share a micro-homologous sequence with the flanking sequence are thought to be scars of imperfect repair of double-strand breaks by alternative endjoining processe.
- To extract IDs, a global analysis of 2,780 cancers of multiple tissue types was performed on indels classified according to a set of 83 channels. Seventeen IDs were reported
- ID6 represented a microhomology-mediated deletion signature seen in *BRCA1*mutated and *BRCA2*mutated cancers described previously.
- ID1, ID2 and ID7 were repeat-mediated IDs highly elevated in tumours with mutations in the proofreading domains of *POLE* or *POLD1* and/or MMR deficiency.

# Rearrangement signatures

- With use of NMF framework, 32 classification channels were proposed for putative RSs extracted from a local analysis of 560 WGS breast cancers
  - The channels took into account how the rearrangement breakpoints were regionally clustered,
  - the rearrangement type (for example, deletion, tandem duplication (TD), inversion or translocation) and
  - the rearrangement size.
  - Three of the six identified RSs correlated with tumour HRD:
    - cancers with *BRCA1* but not *BRCA2* mutations displayed high numbers of RS3 small TDs (less than 10 kb),
    - cancers with *BRCA1* or *BRCA2* mutations showed a substantial number of RS5 deletions (less than 10 kb).
    - The cause of RS1 long TDs (more than 100 kb), also associated with HRD, was not known.
- The number of RSs was recently extended to 15. The 32-channel classification scheme has also been used to report signatures in liver and ovarian cancer cohorts.
- Using a hierarchical Dirichlet process, the PanCancer Analysis of Whole Genomes (PCAWG) Structural Variation Working Group reported 16 RSs in a global analysis of ~2,559 WGS primary cancers, involving ~150,000 structural variations

# THANK YOU