# RepairSig

## Deconvolution of DNA damage and repair contributions to the mutational landscape of cancer

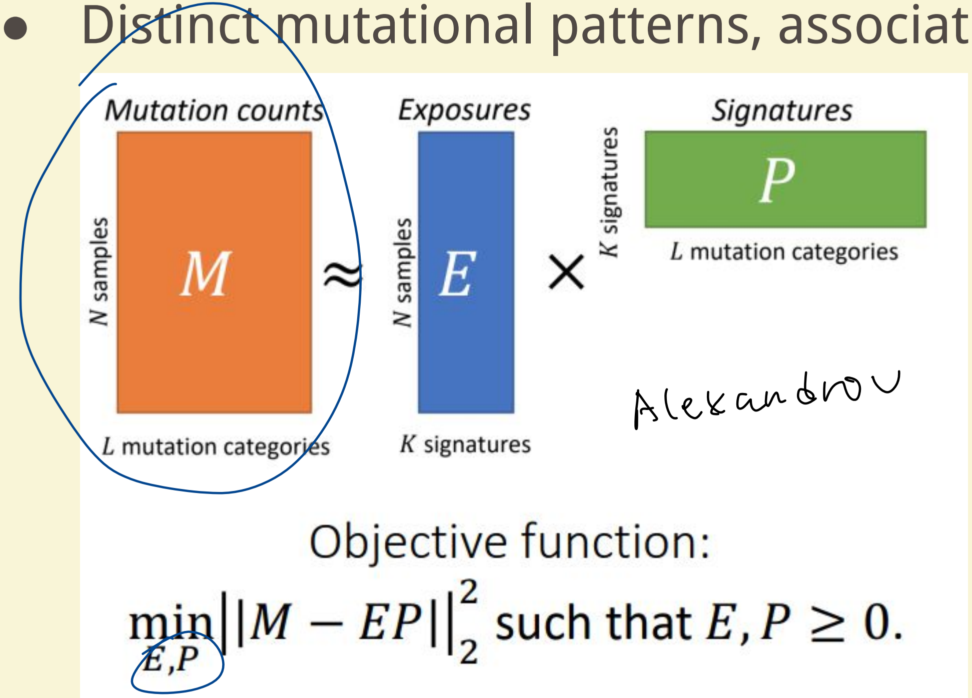
Paper by Wojtowicz, Damian, et al.
Presenter: Baqiao Liu

# Background

- Problem: infer mutational signatures of DNA repair deficiencies
- Current methods: too simple, not capturing non-addictivity
- REPAIRSIG: biologically realistic inference of DNA repair deficiencies

# Mutational Signature

- Distinct mutational patterns, associated with mutational processes



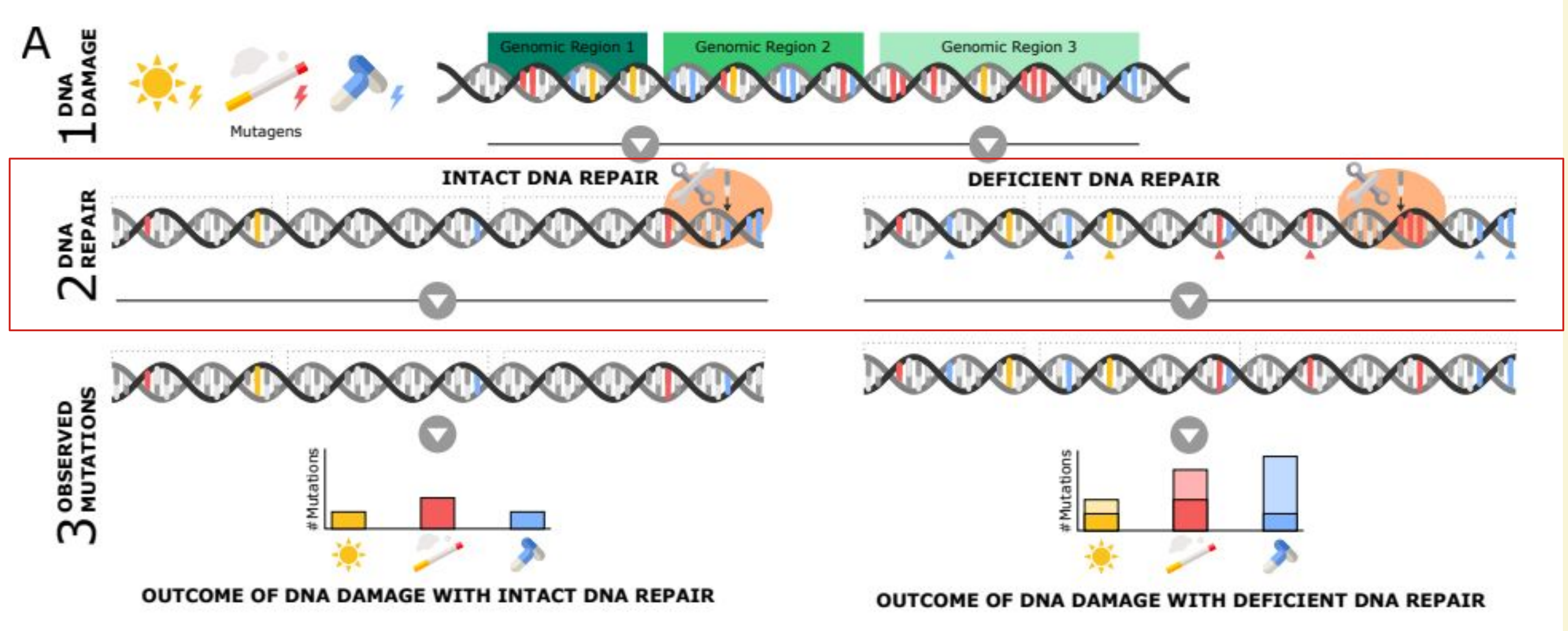
Alexandrov

Objective function:

$$\min_{E,P} ||M - EP||_2^2 \text{ such that } E, P \geq 0.$$

No “secondary” mutatagenatic processes

# DNA Repair Deficiency



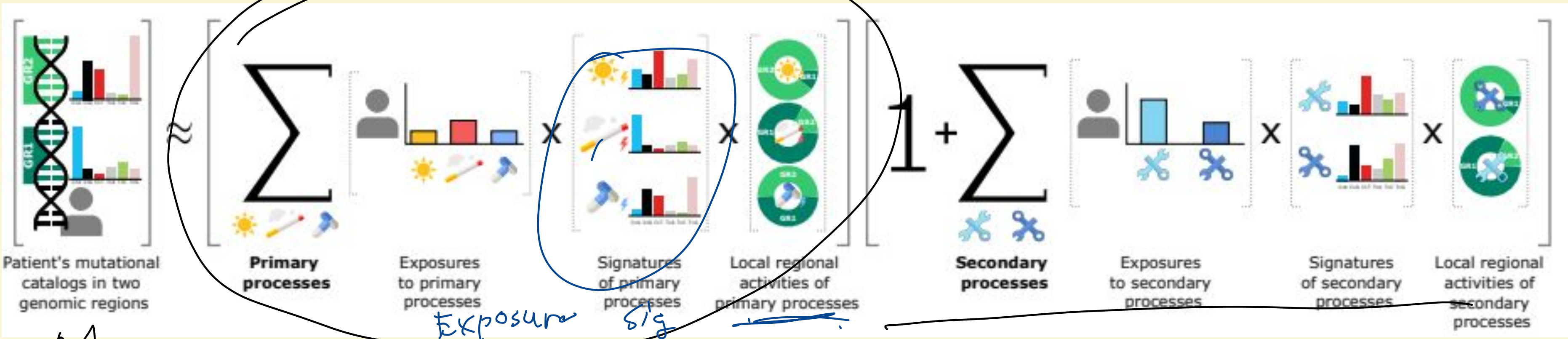
# DNA Repair Deficiency

- Interplay between damage and repair
- Mutation opportunity
  - Distribution of easiness of damage and repair (within genome, across cancer genomes)

# DNA Repair Deficiency

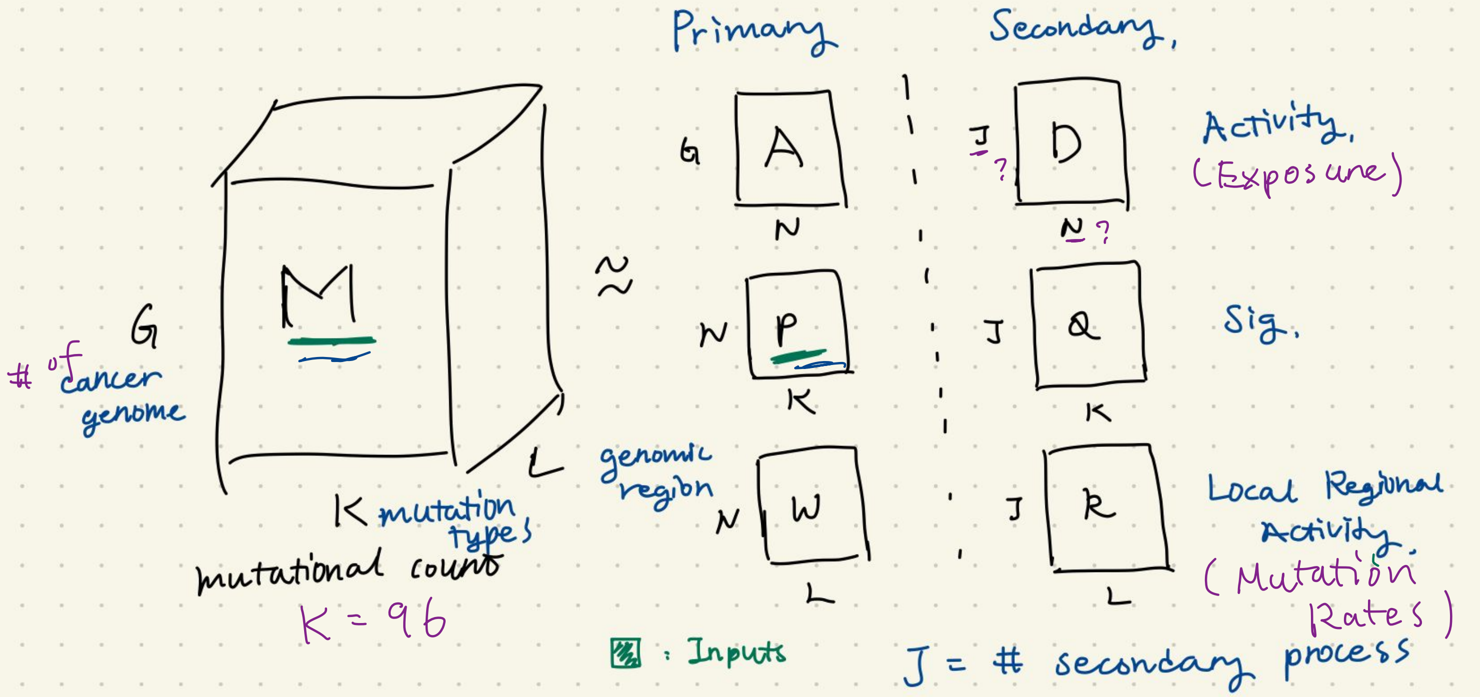
- Problem: infer mutational signatures of DNA repair deficiencies
- Current methods: too simple, not capturing non-addictivity
- RepairSig: biologically realistic inference of DNA repair deficiencies

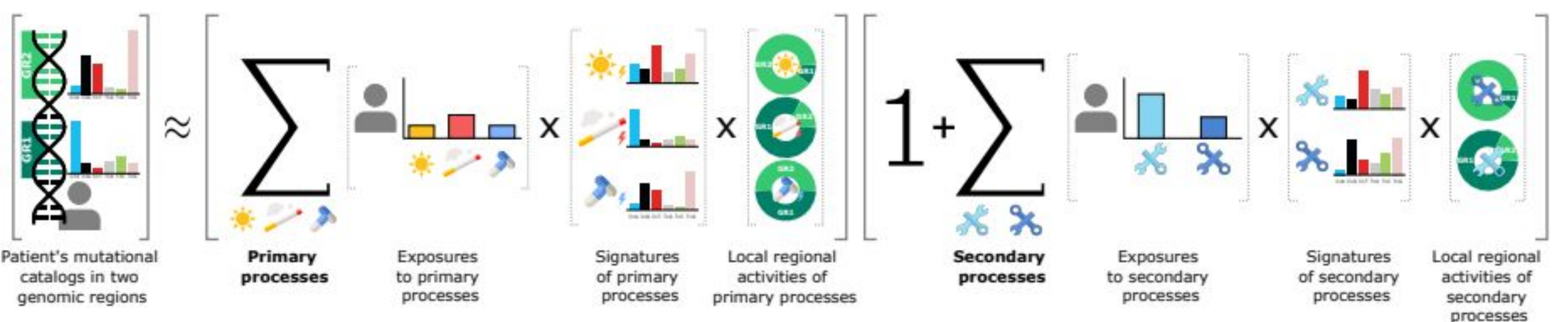
# RepairSig's Model



Exposure Sig

M.


Primary
Secondary,
G
A
N
J ?
D
N ?
Activity,
(Exposure)
N
P
K
J
Q
K
Sig.
genomic region
N
W
L
J
R
L
Local Regional Activity.
(Mutation Rates)
G
# of cancer genome
M
L
≈
K mutation types
mutational count
K = 96
▨ : Inputs
J = # secondary process


GR2
GR1
≈
Σ
x
x
1 + Σ
x
x
GR1
GR2
Patient's mutational catalogs in two genomic regions
Primary processes
Exposures to primary processes
Signatures of primary processes
Local regional activities of primary processes
Secondary processes
Exposures to secondary processes
Signatures of secondary processes
Local regional activities of secondary processes

# RepairSig's Optimization Problem

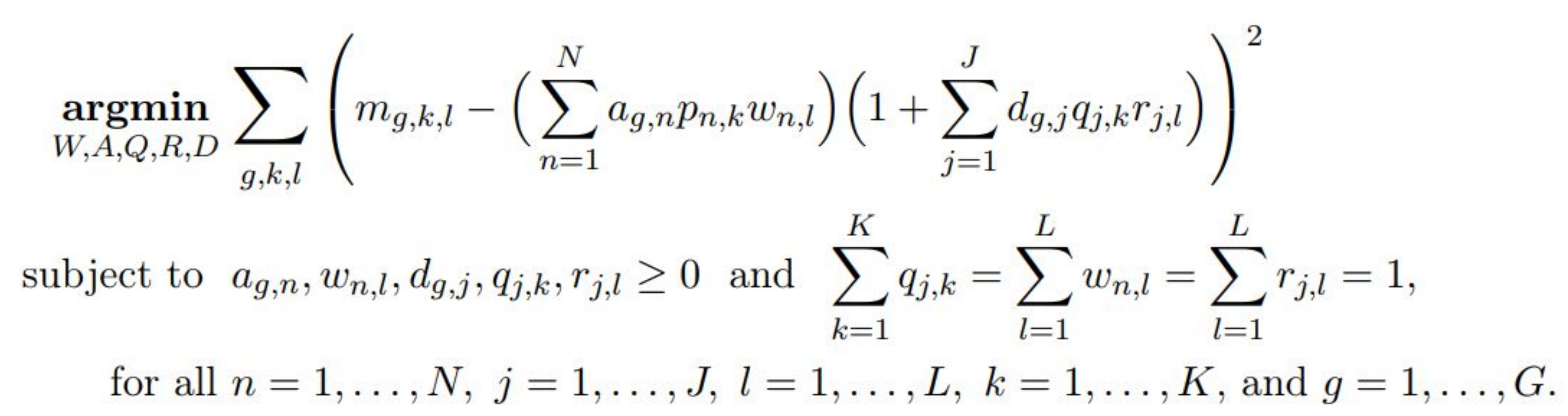
- Find P, W, D, Q, R that minimize norm against approximated M

$$m_{g,k,l} \approx \overbrace{\left(\sum_{n=1}^{N} a_{g,n} p_{n,k} w_{n,l}\right)}^{\text{primary}} \overbrace{\left(\Pi_{j=1}^{J} (1 + d_{g,j} q_{j,k} r_{j,l})\right)}^{\text{secondary}}$$

$$\approx \left(\sum_{n=1}^{N} a_{g,n} p_{n,k} w_{n,l}\right)\left(1 + \sum_{j=1}^{J} d_{g,j} q_{j,k} r_{j,l}\right) = \widehat{m}_{g,k,l}$$

# RepairSig's Optimization Problem

- Find ~~P~~, W, D, Q, R that minimize norm against approximated M
- (Implementation: tensor algebra, gradient descent)



$$\underset{W,A,Q,R,D}{\mathbf{argmin}} \sum_{g,k,l} \left( m_{g,k,l} - \Big(\sum_{n=1}^{N} a_{g,n} p_{n,k} w_{n,l}\Big)\Big(1 + \sum_{j=1}^{J} d_{g,j} q_{j,k} r_{j,l}\Big)\right)^2$$

$$\text{subject to } \; a_{g,n}, w_{n,l}, d_{g,j}, q_{j,k}, r_{j,l} \geq 0 \;\text{ and }\; \sum_{k=1}^{K} q_{j,k} = \sum_{l=1}^{L} w_{n,l} = \sum_{l=1}^{L} r_{j,l} = 1,$$

$$\text{for all } n = 1, \ldots, N,\; j = 1, \ldots, J,\; l = 1, \ldots, L,\; k = 1, \ldots, K, \text{ and } g = 1, \ldots, G.$$

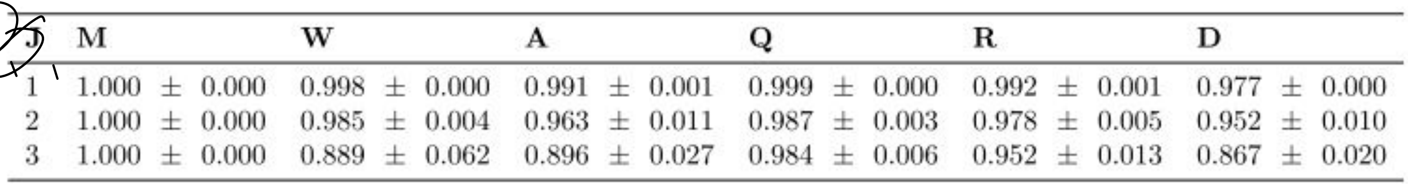
# Results

- Simulation
- BRCA data
- Cross-cancer analysis (not shown here)

# Results: Simulation

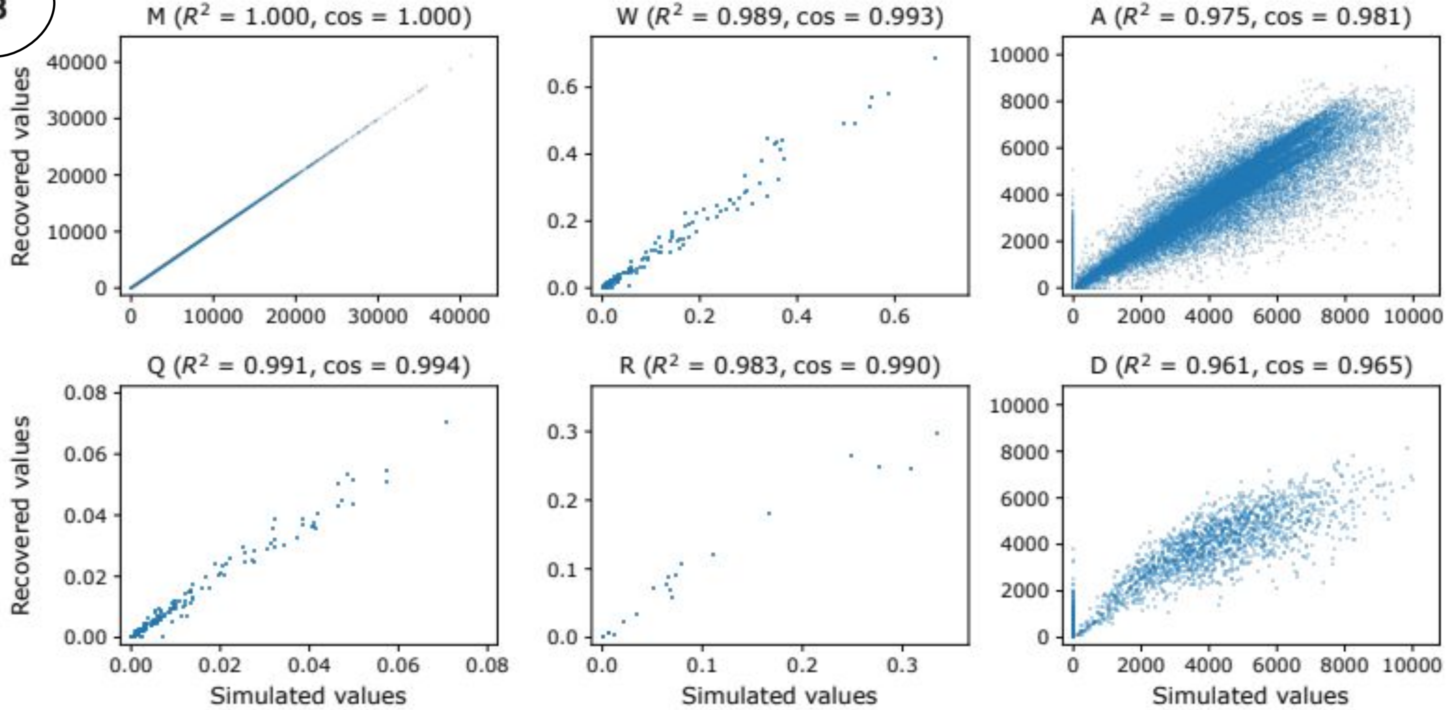
- P, Q from real data. A, D, W, R drawn from distribution
- Approximated M constructed from P, Q, A, D, W, R
- The correlation coefficient is shown

**A**



| J | M | W | A | Q | R | D |
|---|---|---|---|---|---|---|
| 1 | 1.000 ± 0.000 | 0.998 ± 0.000 | 0.991 ± 0.001 | 0.999 ± 0.000 | 0.992 ± 0.001 | 0.977 ± 0.000 |
| 2 | 1.000 ± 0.000 | 0.985 ± 0.004 | 0.963 ± 0.011 | 0.987 ± 0.003 | 0.978 ± 0.005 | 0.952 ± 0.010 |
| 3 | 1.000 ± 0.000 | 0.889 ± 0.062 | 0.896 ± 0.027 | 0.984 ± 0.006 | 0.952 ± 0.013 | 0.867 ± 0.020 |

**B**

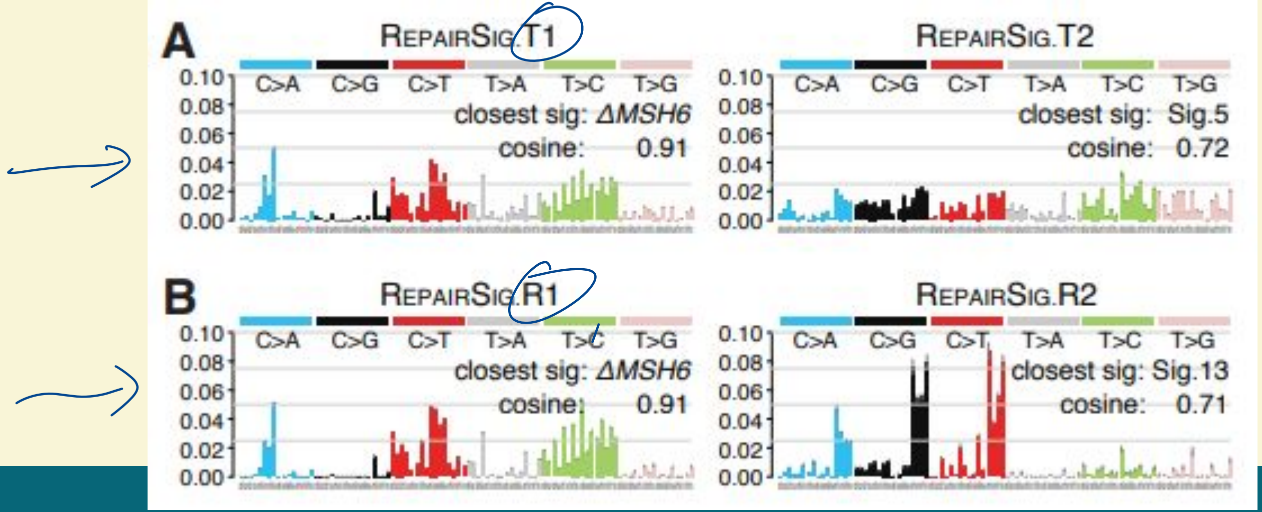


# Results: BRCA

- Previous work: >= 3 signatures associated with MMR deficiency (MMRd)
    - MMR: mismatch repair, correcting non-complementary nucleotides
    - DNA repair
- Hypothesis: these signatures are not needed (as primary signatures). They should now be explicitly modeled as secondary signatures

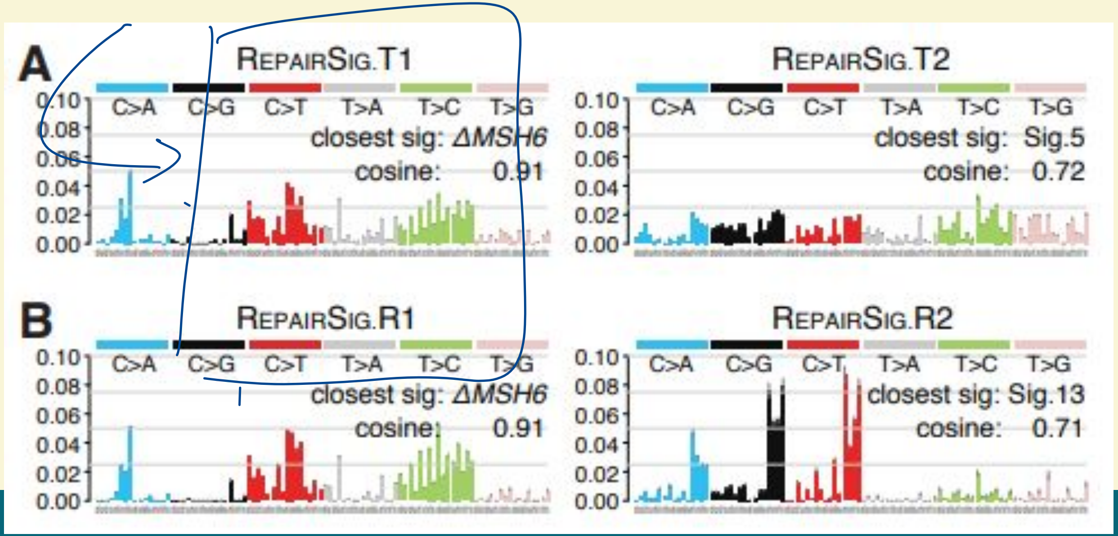
# Results: BRCA

- Two settings (different partitioning of genome into local regions)
    - strand-specific direction of gene transcription
    - discretized replication timing data
- Four signatures (two each from each setting), two the same (T1, R1)



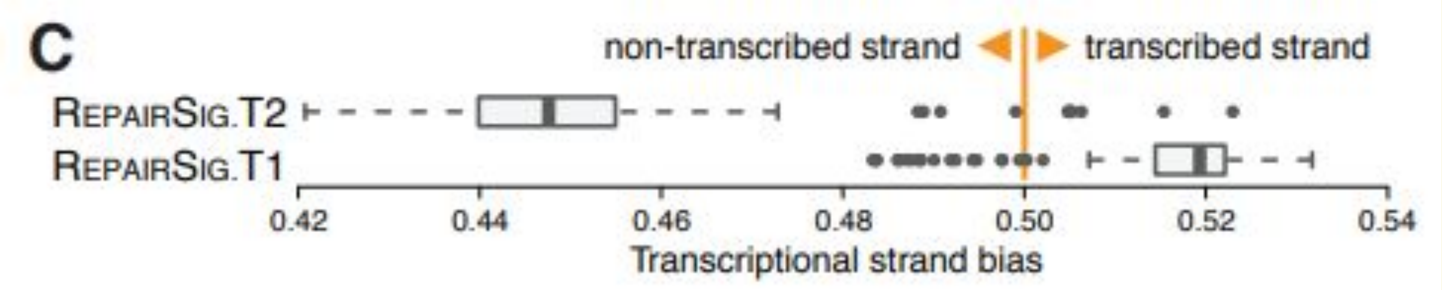
# Results: BRCA, Signatures

- T1 = R1 related to MMR deficiency
- Others not related to known MMR deficiency signatures
    - Argued to not be related to MMRd
    - One secondary signature of MMRd replacing three (primary) MMRd



# Results: BRCA, Signatures

- T2 Signature
    - High transcriptional strand bias
    - Possibly related to TC-NER “repairs the transcribed strand of transcriptionally active genes more efficiently than nontranscribed strand and transcriptionally silent DNA”
    - Most related COSMIC signature (5) is related to DNA repair



# Results: BRCA, Signatures

- R2 Signature
    - Most similar to COSMIC Signatures 2 and 13 (0.71, 0.67, cosine similarity)
    - These two signatures are included as primary signatures in input
    - Implying that R2 is not simply linear combination of 2 and 13
    - Theorized to be more active than Signatures 2 and 13 in early replicating regions

# Acknowledgements

- Figures stolen from:
  - Original paper (preprint, bioRxiv)
  - Dr. El-Kebir's slides
- Some figures from Dr. El-Kebir