Notes on gender associations in indel signatures, 2025 01 20

See file investigate\_gender\_bias.R

Technical errors are possible, suggest to sanity check.

Quite a few samples had no “Gender”. I removed these. See lines 25..32.

Was this expected or is there an upstream data-wrangling issue?

There were strong associations for cancer type Other. Presumably due to differences in prevalences of subtypes of “Other” between men and women

I removed signature, cancertype pairs where the signature was rare or almost always present.

Lines 91.95

Only Lung and Skin had differences with FDR < 0.2. Lines 131…155.

In Lung, ID3 is enriched in men, probably reflects more lung cancer cases in men due to smoking than lung cancer cases in women.

This is an example of how gender differences in signatures might reflect gender differences in exposure.

In Skin, C\_ID13 is enriched in men. Any idea why? Different kinds of UV exposures?

Filtering just by p value

lines 176..182

Bladder/C\_ID19 was high in men. Exposure?

Bone.SoftTissue/C\_ID8 was high in men. Differences in cancer subtype prevalence?

Kidney/C\_ID19 was high in men – subtype? Exposure?

Liver/H\_ID25 high in men – exposure?

Lines 188..195

Head high in women; subtype prevalence?

Lymphoid high in women; subtype prevalence