Aim and deliverables for the functional impact paper

Decipher overall functional burdening in cancer genomes in the PCAWG project.

- Avg cancer has ~5 drivers & ~5000 mutations. What is the overall burdening of the many passengers in different cancers ?

- Look at Overall variation burden observed in various genomic elements (coding & noncoding) in different PCAWG cohorts.
  - ➢ Comparison between real and simulated data to highlight genomic elements with significant burden from passengers in different cohorts
  - ➢ This work will provide **comprehensive functional annotations across all of pcawg** (FunSeq & aloft score)

- Coding and noncoding functional impact score distribution across pan-cancer cohorts.
  - ➢ Enrichment/depletion of high impact passengers (other than drivers) in gene block/neighborhood
  - ➢ Correlation of passenger burdening with downstream gene expression changes
  - ➢ Framework to evaluate structural variation impact score

- Comparison between somatic and germline variation burdening
  - ➢ Investigate influence of germline mutational burden on the somatic genome variation profile

- Decipher the the differential passenger burdening in various cohorts (how it relates to mechanism)
  - ➢ Relate to different Signature, Ageing, sub-clonality & other clinical information

Additive effects model to quantify cumulative variance

- Model for the effect of an individual SNP on a phenotype

$$y_j = \mu + x_{ij}a_i + e_j$$

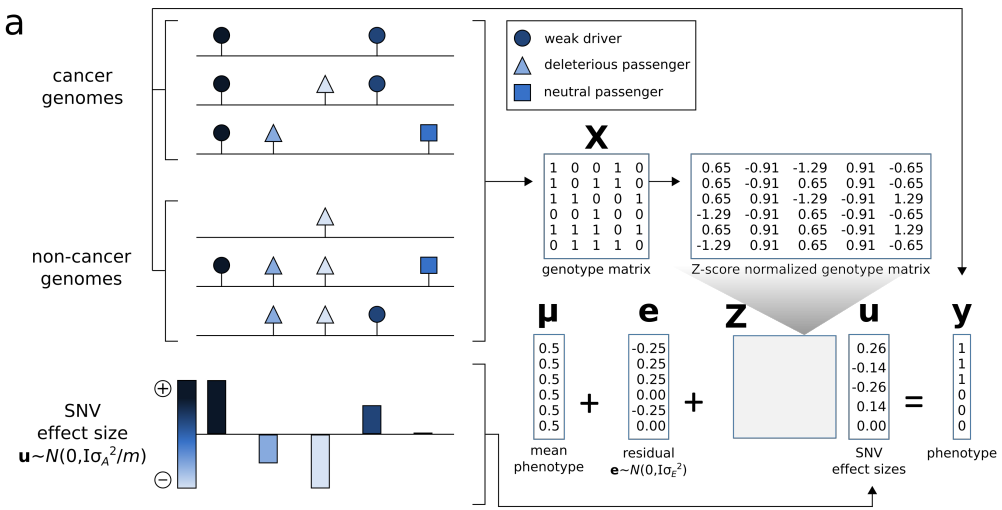
where: y=phenotype; x_ij is the 'genetic dosage' of the i'th SNP in individual j, taking values {0,1}; a_i is the fixed effect size of SNP i, and e_j is the residual effect

- Extension to model the combined effects of multiple SNPs

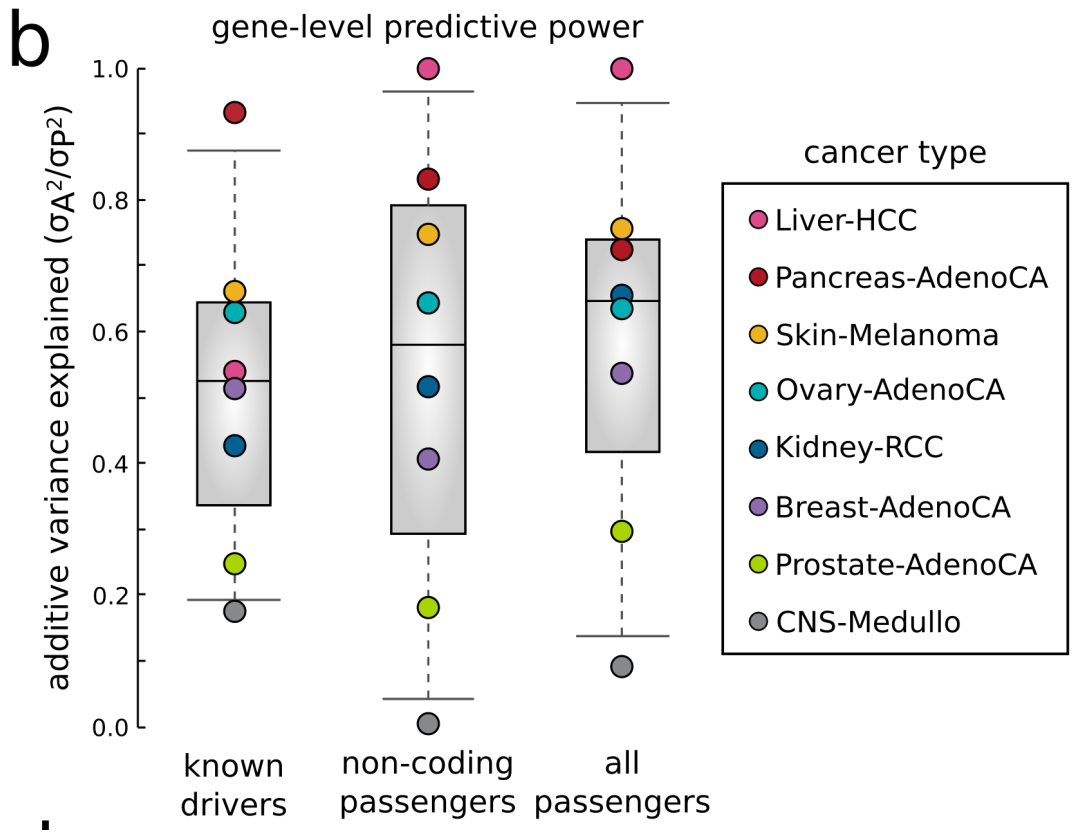
$$y_j = \mu + g_j + e_j \text{ and } g_j = \sum_{i=1}^{m} z_{ij}u_i$$

$$g_j \sim N(0, \sigma_g^2 = m\sigma_u^2) \qquad \mathbf{u} \sim N(\mathbf{0}, \mathbf{I}\sigma_u^2)$$

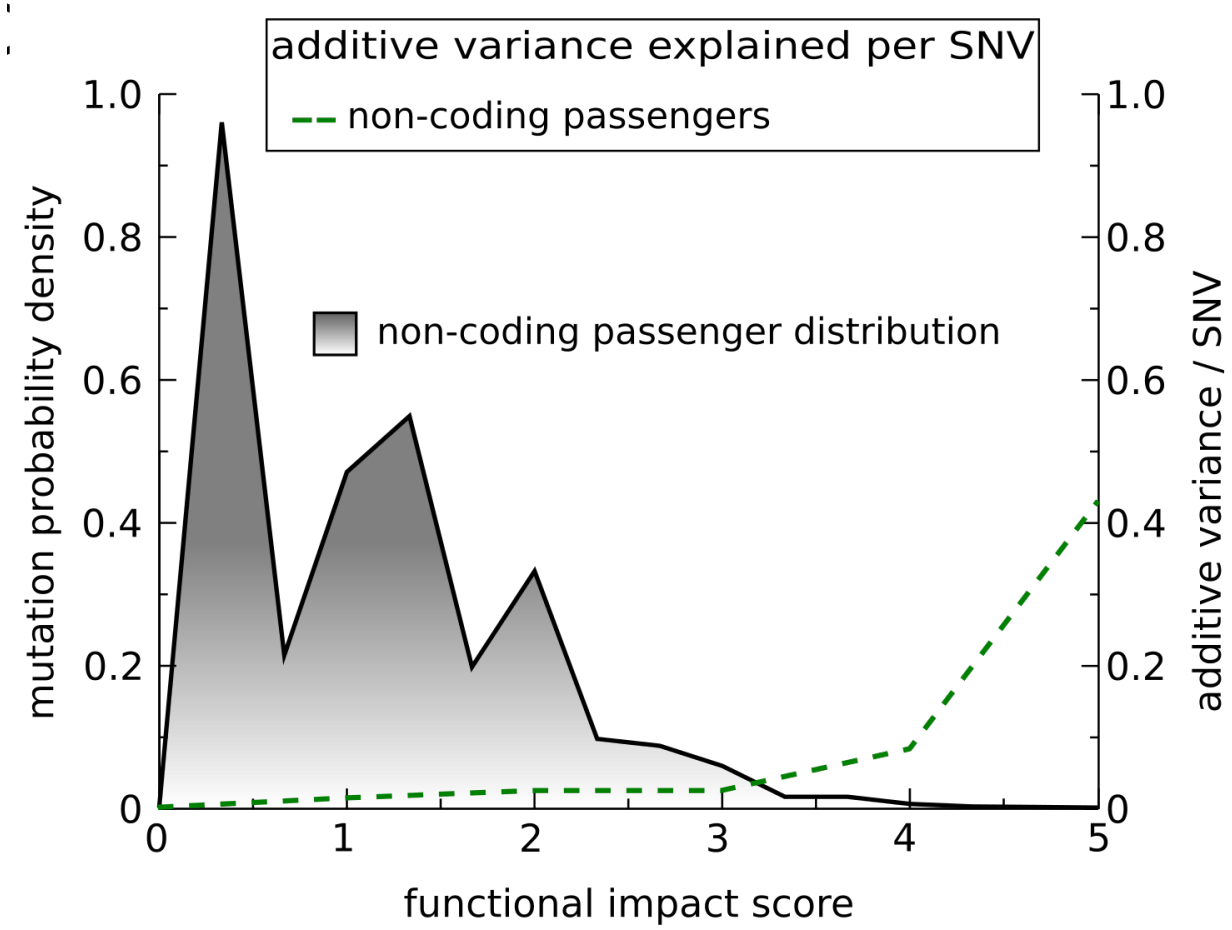
where: z_ij is a 'normalized genetic dosage', i.e. the z-score of x_ij; u_i is the effect size of SNP i treated as a random variable; g_j is the combined effect of all SNPs for individual j



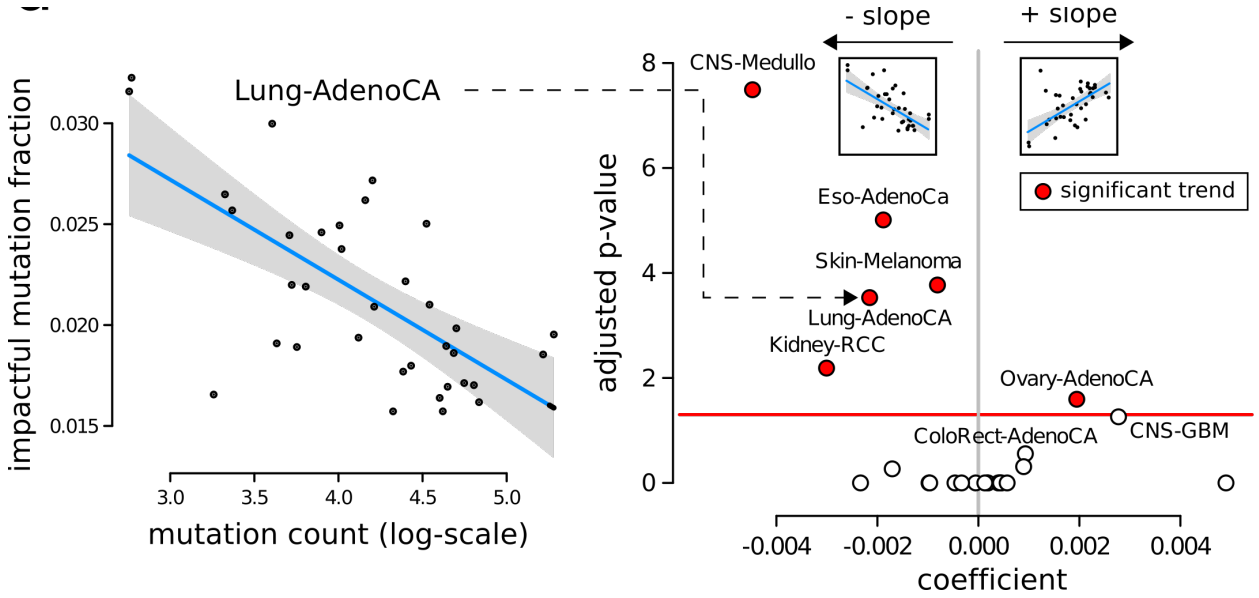
Additive variance for multiple cancer cohorts in PCAWG



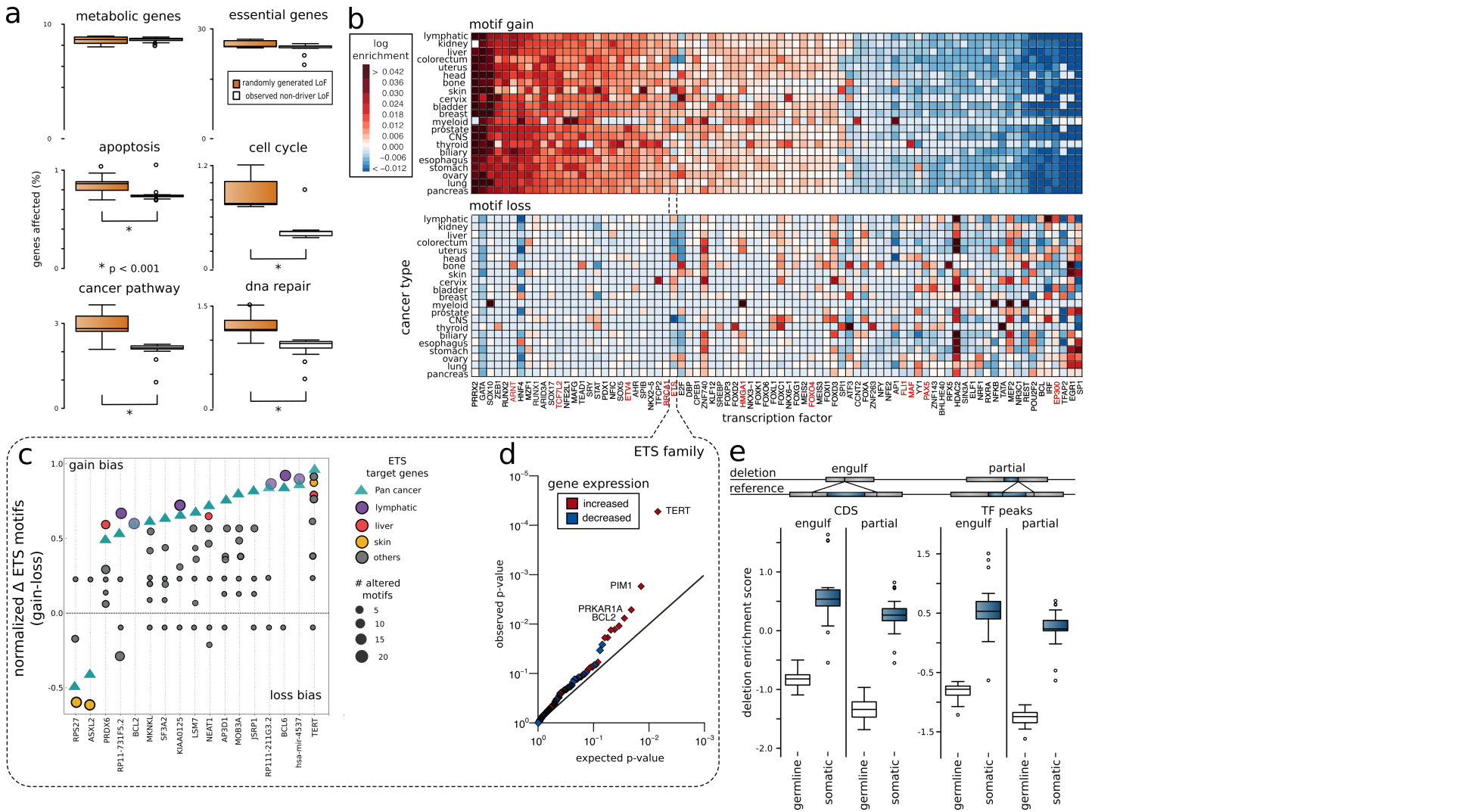
Non-coding predicted functional impact score distribution



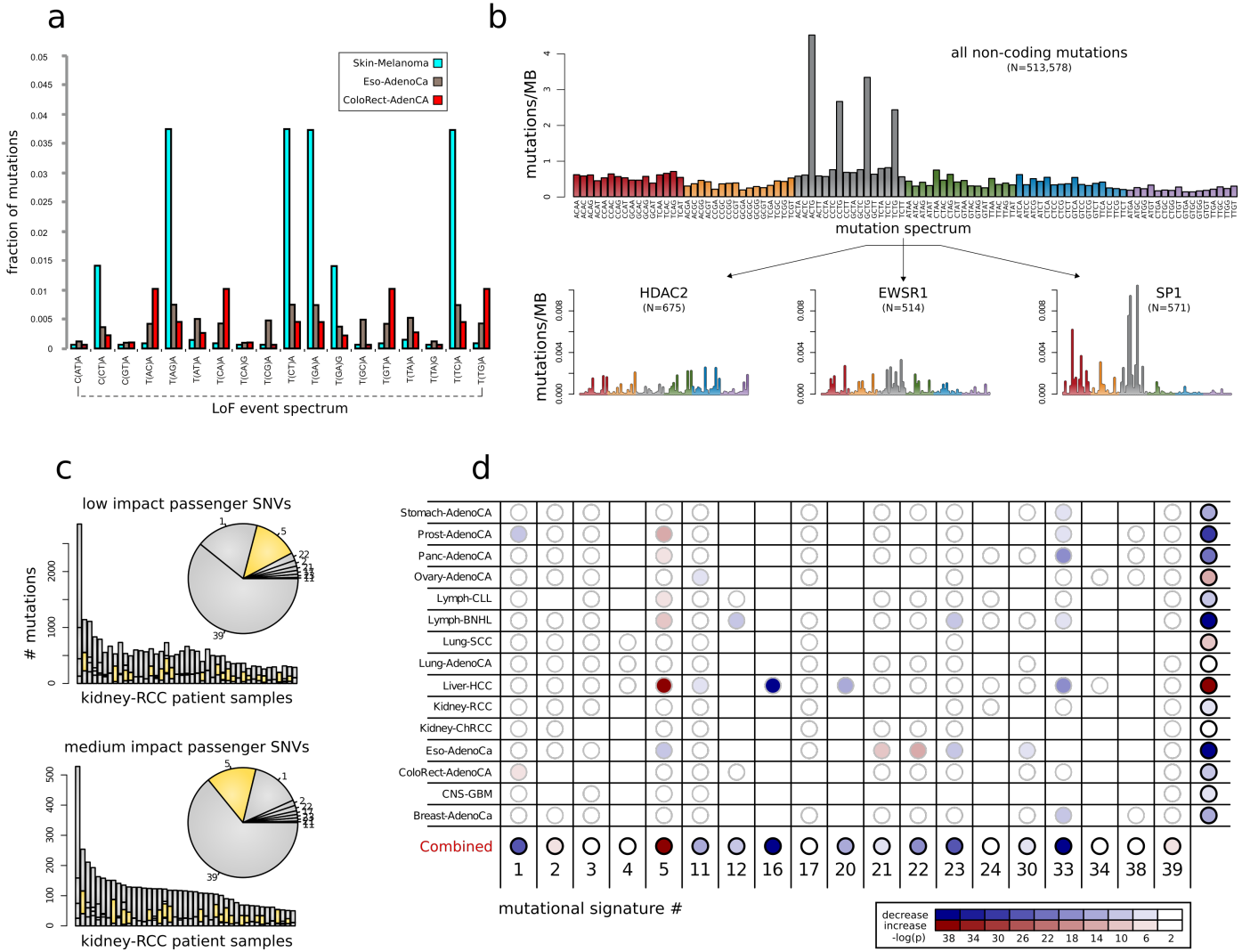
Fraction of high impact nominal passengers in PCAWG cohorts



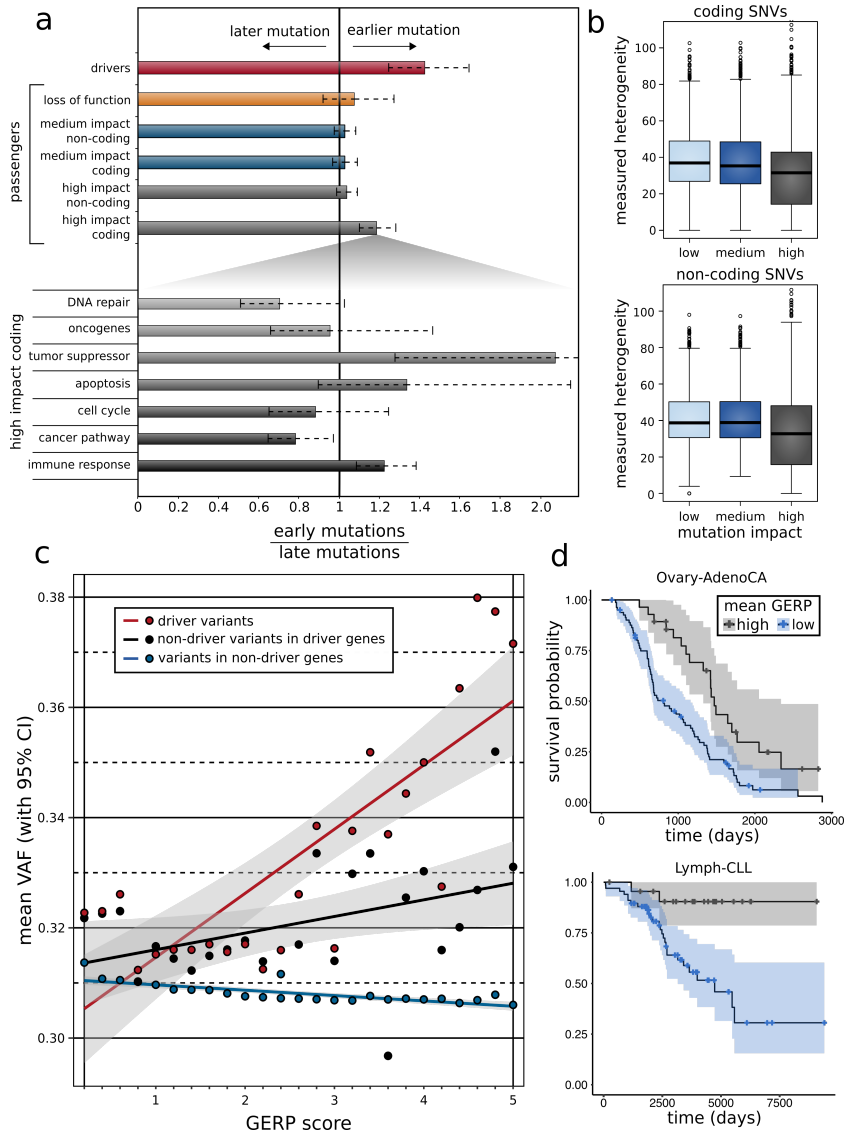
Alterations in TF binding landscape: overall burden



De-convoluting mutational signature and functional impact of nominal passengers in PCAWG



Subclonal architecture of nominal passengers in PCAWG



Extending canonical model of driver and passengers