Epistatic effects among lung adenocarcinoma somatic mutations across oncogenesis

Jorge Alfaro-Murillo, Krishna Dasari & Jeffrey Townsend

Biostatistics Department

Yale SCHOOL OF PUBLIC HEALTH

Motivation

- *KRAS* and *TP53* are considered early drivers of LUAD but which is first?
- Smoking causes LUAD but is it just because of more mutations?
- Why is *EGFR* in so many non-smoker LUAD cases?

Motivation

- *KRAS* and *TP53* are considered early drivers of LUAD but which is first?
- Smoking causes LUAD but is it just because of more mutations?
- Why is *EGFR* in so many non-smoker LUAD cases?

Motivation

- *KRAS* and *TP53* are considered early drivers of LUAD but which is first?
- Smoking causes LUAD but is it just because of more mutations?
- Why is *EGFR* in so many non-smoker LUAD cases?

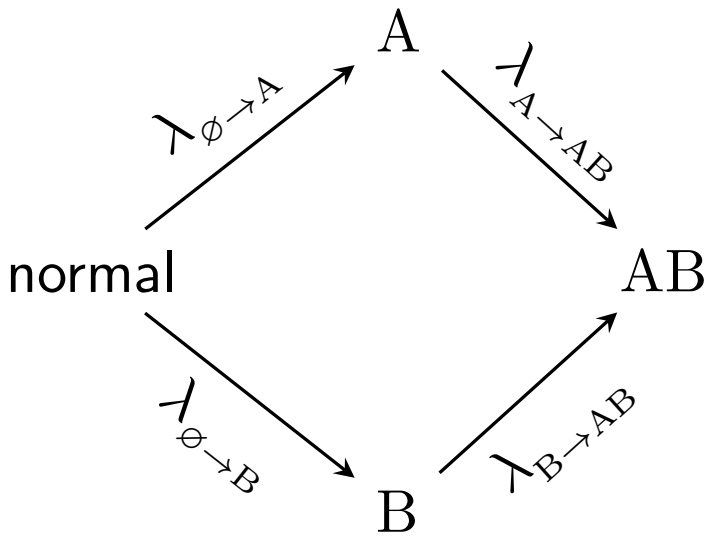
Model

- There is a constant rate at which mutations occur and are selected to high frequency (flux)
- The flux depends on the gene that will mutate as well as the current somatic genotype (previous mutations)

- Epistasis occurs when $\lambda_{\emptyset \to \mathrm{A}} \neq \lambda_{B \to \mathrm{AB}}$
- Given the number of tumors in each genotype we can estimate the fluxes

Model

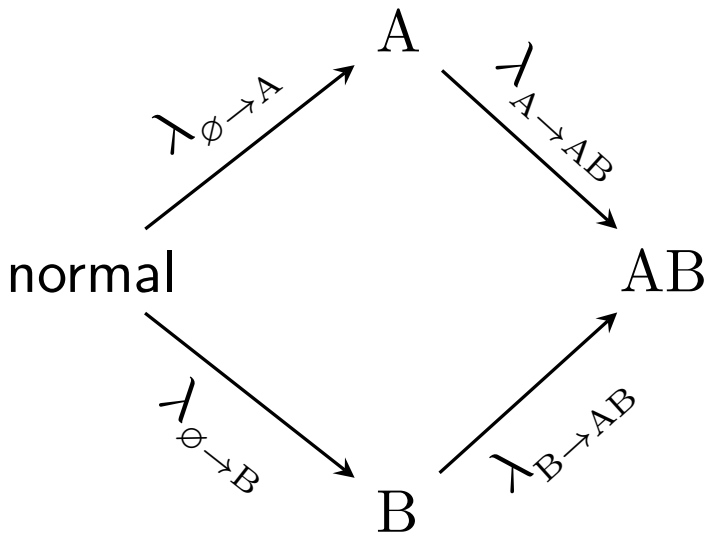
- There is a constant rate at which mutations occur and are selected to high frequency (flux)
- The flux depends on the gene that will mutate as well as the current somatic genotype (previous mutations)



- Epistasis occurs when $\lambda_{\varnothing \to \mathrm{A}} \neq \lambda_{B \to \mathrm{AB}}$
- Given the number of tumors in each genotype we can estimate the fluxes

Model

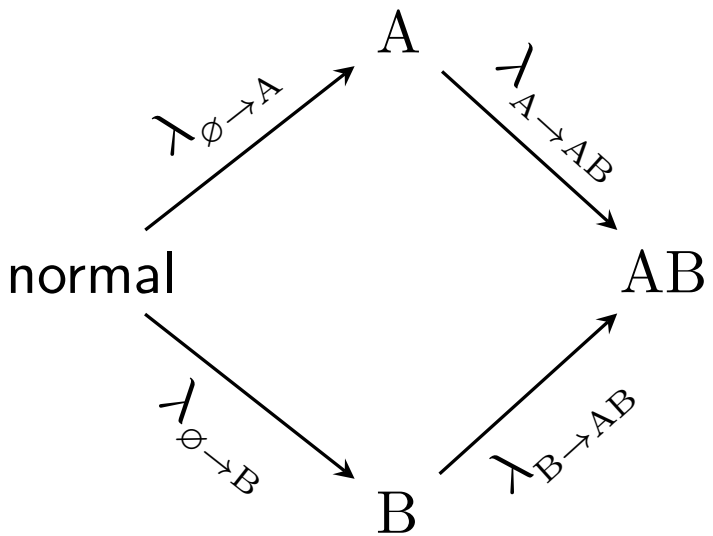
- There is a constant rate at which mutations occur and are selected to high frequency (flux)
- The flux depends on the gene that will mutate as well as the current somatic genotype (previous mutations)



- Epistasis occurs when $\lambda_{\varnothing \to \mathrm{A}} \neq \lambda_{B \to \mathrm{AB}}$
- Given the number of tumors in each genotype we can estimate the fluxes

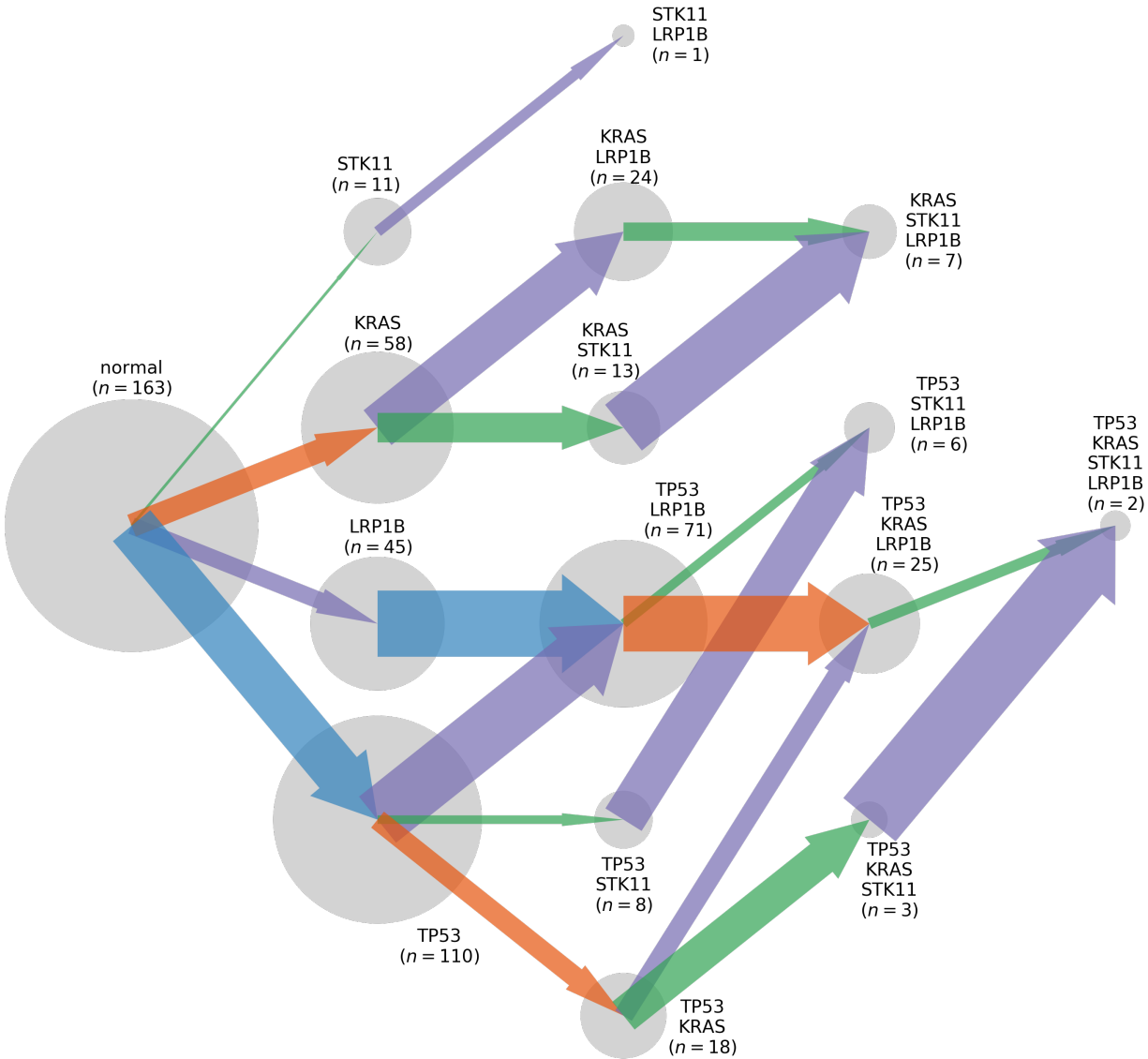
Model

- There is a constant rate at which mutations occur and are selected to high frequency (flux)
- The flux depends on the gene that will mutate as well as the current somatic genotype (previous mutations)



- Epistasis occurs when $\lambda_{\varnothing \to \mathrm{A}} \neq \lambda_{B \to \mathrm{AB}}$
- Given the number of tumors in each genotype we can estimate the fluxes

Epistasis in most commonly mutated genes in TCGA



Deconvolution of mutation and selection

- The flux is deconvolved into the mutation rate of the gene and the strength of selection on mutations of gene
- Mutation rate are obtained for each possible variant of each gene considering:
  - Molecular signatures
  - Gene expression
  - Chromatin marks
  - Replication times

Deconvolution of mutation and selection

- The flux is deconvolved into the mutation rate of the gene and the strength of selection on mutations of gene
- Mutation rate are obtained for each possible variant of each gene considering:
  - Molecular signatures
  - Gene expression
  - Chromatin marks
  - Replication times

Deconvolution of mutation and selection

- The flux is deconvolved into the mutation rate of the gene and the strength of selection on mutations of gene
- Mutation rate are obtained for each possible variant of each gene considering:
  - Molecular signatures
  - Gene expression
  - Chromatin marks
  - Replication times

Deconvolution of mutation and selection

- The flux is deconvolved into the mutation rate of the gene and the strength of selection on mutations of gene
- Mutation rate are obtained for each possible variant of each gene considering:
  - Molecular signatures
  - Gene expression
  - Chromatin marks
  - Replication times

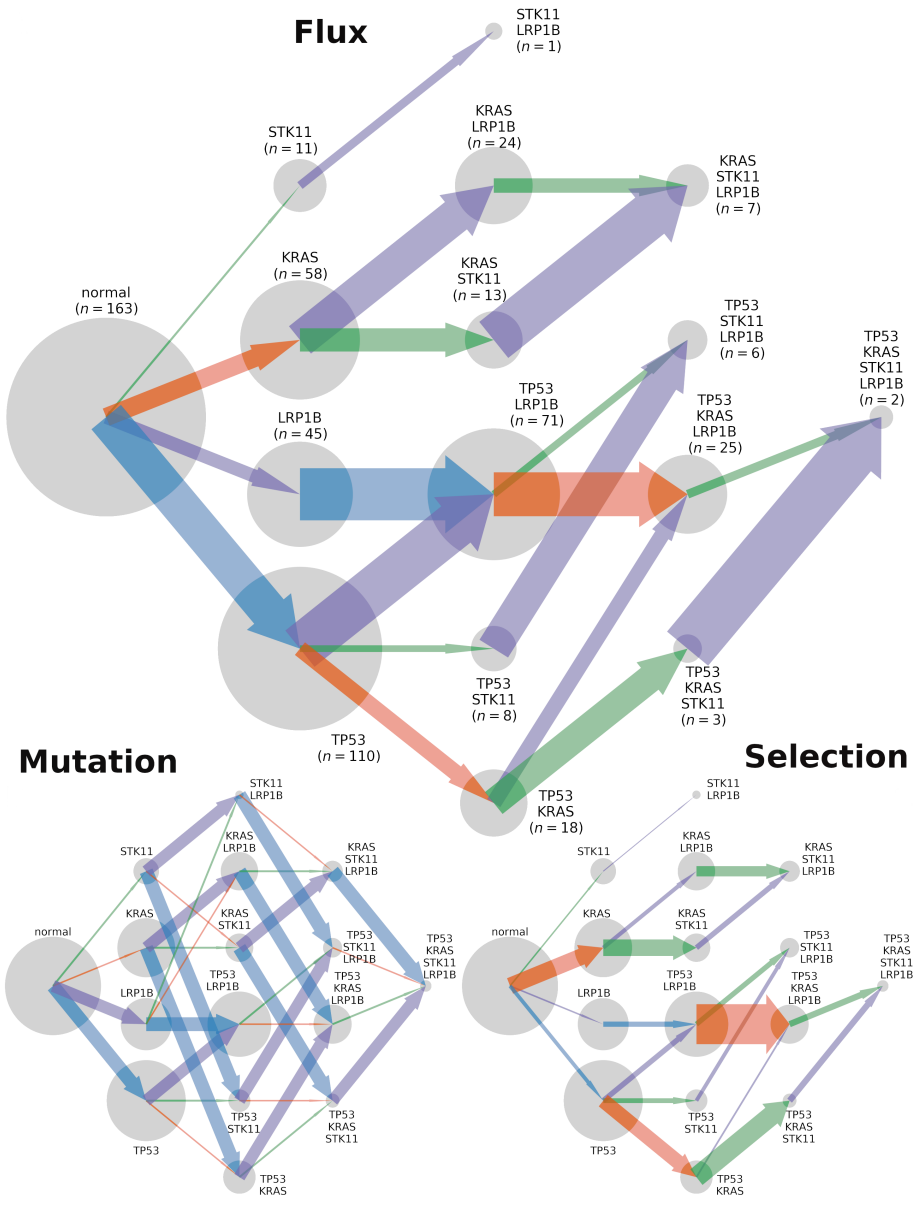
Deconvolution of mutation and selection

- The flux is deconvolved into the mutation rate of the gene and the strength of selection on mutations of gene
- Mutation rate are obtained for each possible variant of each gene considering:
  - Molecular signatures
  - Gene expression
  - Chromatin marks
  - Replication times

Deconvolution of mutation and selection

- The flux is deconvolved into the mutation rate of the gene and the strength of selection on mutations of gene
- Mutation rate are obtained for each possible variant of each gene considering:
  - Molecular signatures
  - Gene expression
  - Chromatin marks
  - Replication times

Deconvolution of mutation and selection



Larger data set

- How about other genes? What is the effect of smoking?
- Aggregate multiple data
  1. The Cancer Genome Atlas (TCGA)
  2. AACR Project GENIE
  3. Kenfield et al. Tob Control, 2008
  4. Chen et al. Nat Genet, 2020
  5. Rizvi et al. Science, 2015
  6. Hellmann et al. Cancer Cell, 2018
  7. Jamal-Hanjani et al. N Engl J Med, 2017
  8. Abbosh et al. Nature, 2017
  9. Imielinski et al. Cell, 2012
  10. Ding et al. Nature, 2008
  11. Jordan et al. Cancer Discov, 2017
- Total: 8,487 non-metastatic LUAD samples
- Classified 1,073 smokers and 447 non-smokers with clinical data and COSMIC SBS4

Larger data set

- How about other genes? What is the effect of smoking?
- Aggregate multiple data
  1. The Cancer Genome Atlas (TCGA)
  2. AACR Project GENIE
  3. Kenfield et al. Tob Control, 2008
  4. Chen et al. Nat Genet, 2020
  5. Rizvi et al. Science, 2015
  6. Hellmann et al. Cancer Cell, 2018
  7. Jamal-Hanjani et al. N Engl J Med, 2017
  8. Abbosh et al. Nature, 2017
  9. Imielinski et al. Cell, 2012
  10. Ding et al. Nature, 2008
  11. Jordan et al. Cancer Discov, 2017
- Total: 8,487 non-metastatic LUAD samples
- Classified 1,073 smokers and 447 non-smokers with clinical data and COSMIC SBS4

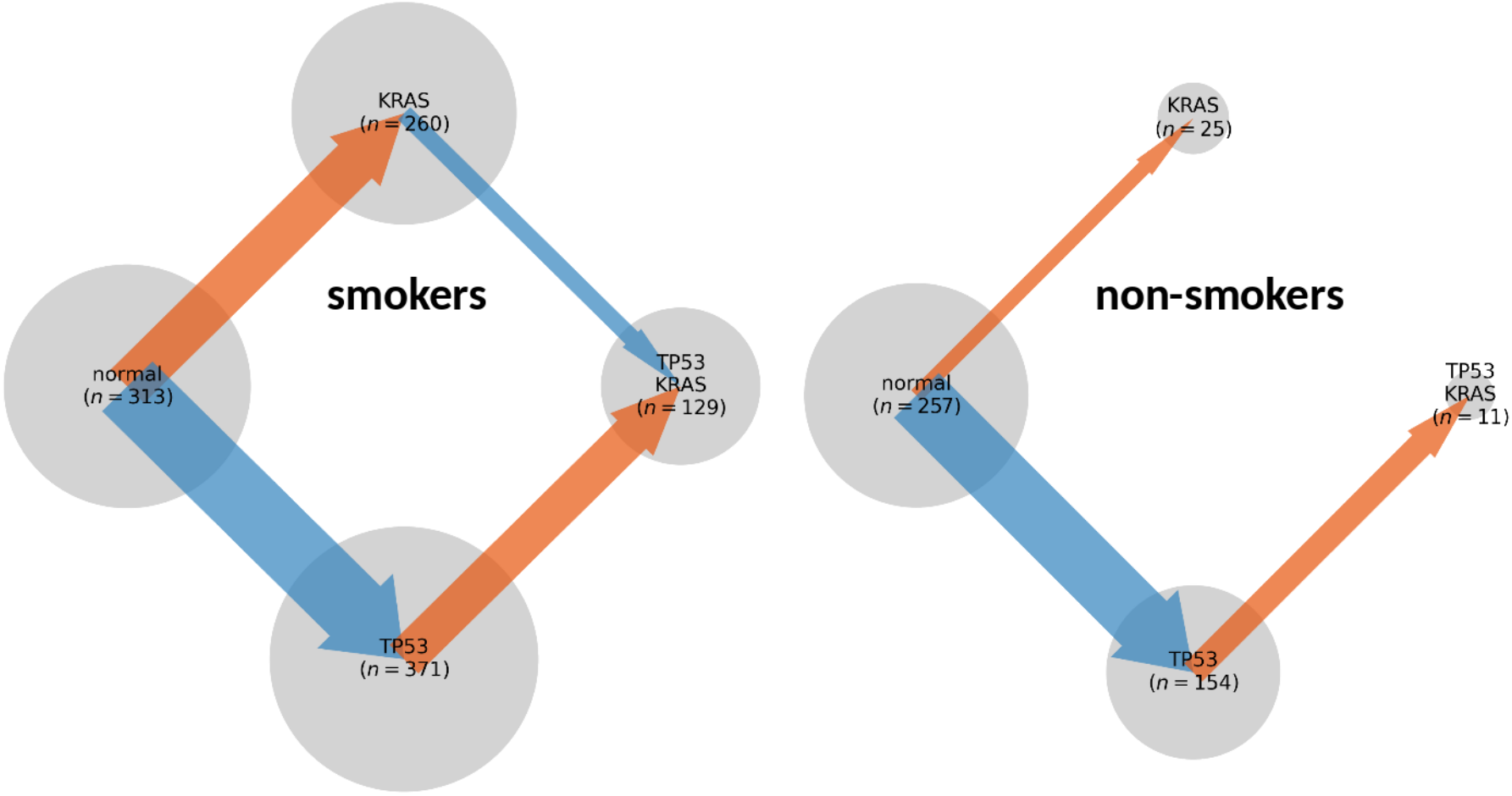
Larger data set

- How about other genes? What is the effect of smoking?
- Aggregate multiple data
  1. The Cancer Genome Atlas (TCGA)
  2. AACR Project GENIE
  3. Kenfield et al. Tob Control, 2008
  4. Chen et al. Nat Genet, 2020
  5. Rizvi et al. Science, 2015
  6. Hellmann et al. Cancer Cell, 2018
  7. Jamal-Hanjani et al. N Engl J Med, 2017
  8. Abbosh et al. Nature, 2017
  9. Imielinski et al. Cell, 2012
  10. Ding et al. Nature, 2008
  11. Jordan et al. Cancer Discov, 2017
- Total: 8,487 non-metastatic LUAD samples
- Classified 1,073 smokers and 447 non-smokers with clinical data and COSMIC SBS4

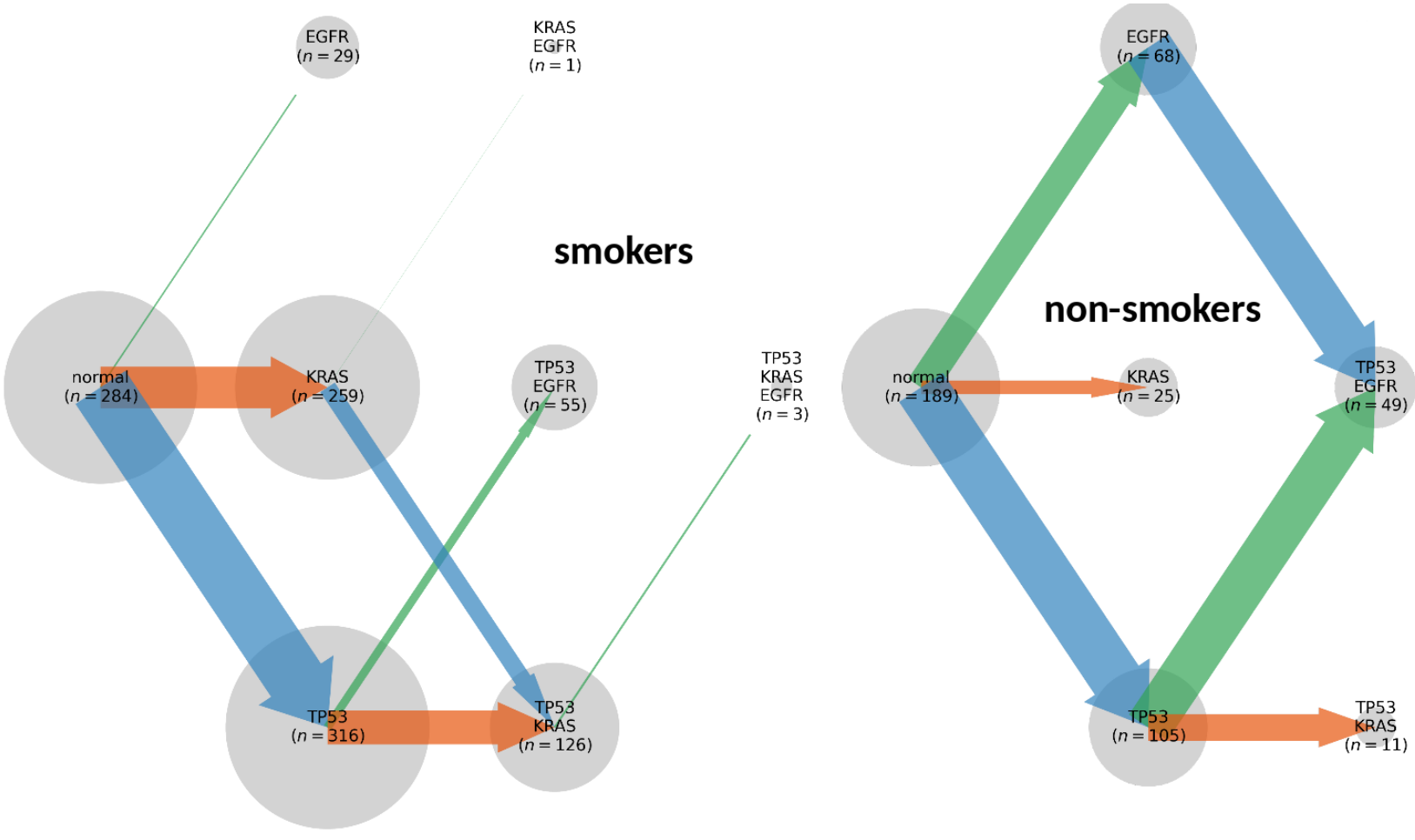
Larger data set

- How about other genes? What is the effect of smoking?
- Aggregate multiple data
  1. The Cancer Genome Atlas (TCGA)
  2. AACR Project GENIE
  3. Kenfield et al. Tob Control, 2008
  4. Chen et al. Nat Genet, 2020
  5. Rizvi et al. Science, 2015
  6. Hellmann et al. Cancer Cell, 2018
  7. Jamal-Hanjani et al. N Engl J Med, 2017
  8. Abbosh et al. Nature, 2017
  9. Imielinski et al. Cell, 2012
  10. Ding et al. Nature, 2008
  11. Jordan et al. Cancer Discov, 2017
- Total: 8,487 non-metastatic LUAD samples
- Classified 1,073 smokers and 447 non-smokers with clinical data and COSMIC SBS4

Epistasis of *TP53* and *KRAS*

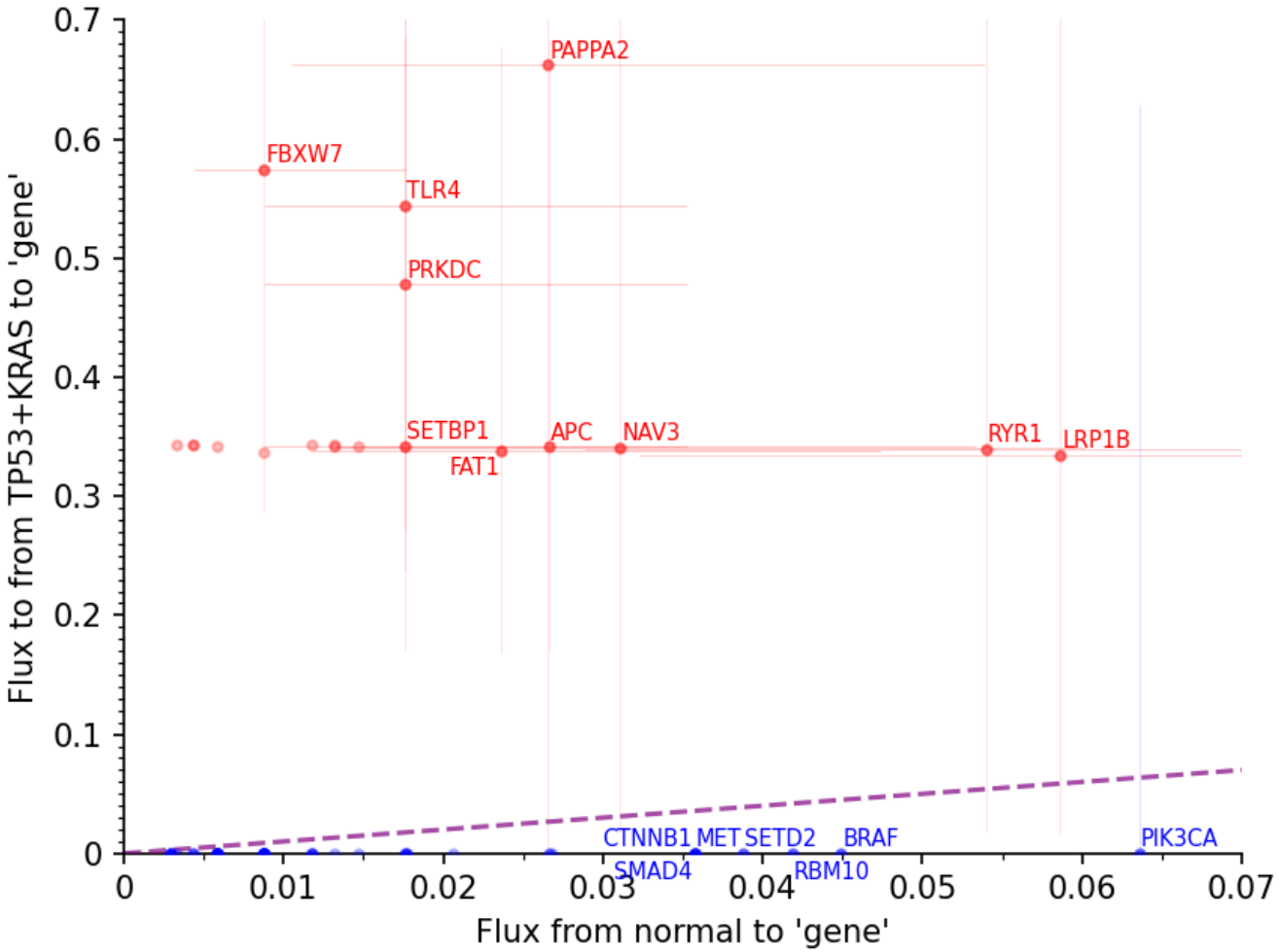


Epistasis of *TP53*, *KRAS* and *EGFR*



Epistasis other genes with *TP53+KRAS*

non-smokers



Epistasis other genes with *TP53+KRAS*

smokers