Kmer identification

De novo search from chm13 reference
(Aganezov et al., *Science*, 2022)

Enhanced Satellite annotations
(Altemose et al., *PLOS Computational Biology*, 2014)

Epigenetic annotations

ENCODE Chip-seq experiments in GM12878
(ENCFF001SUG, ENCFF001SUI, ENCFF001SUJ,
ENCFF001SUE, ENCFF001SUL, ENCFF001SUF,
ENCFF001SUN, ENCFF001SUO, ENCFF001SUP,
ENCFF001SUQ)
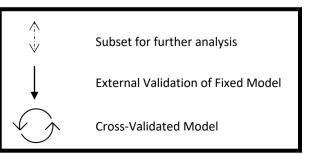(The ENCODE Project Consortium, *Nature*, 2021)

Chromatin states in GM12878
(Ho et al., *Nature*, 2014)
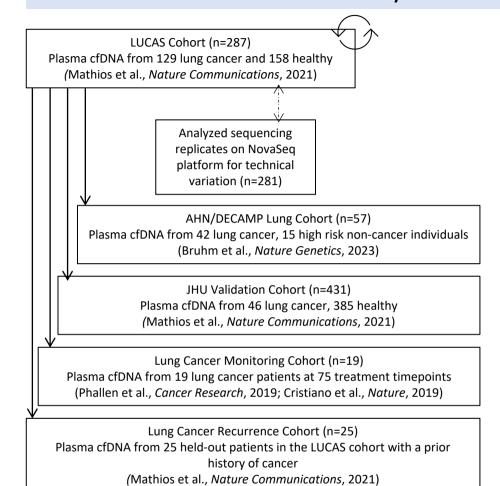
Tissue analyses with ARTEMIS

PCAWG (n=1050)
525 matched tumor and normal pairs from lung (n=172),
liver (n=108), ovarian (n=84), colorectal (n=120), breast
(n=182), bladder (n=46), thyroid (n=96), head and neck
squamous (n=88), prostate (n=38), cervical (n=40), and
gastric (n=76) cancers
(The ICGC/TCGA PCAWG Consortium, *Nature*, 2020)

Plasma analyses with ARTEMIS

LUCAS Cohort (n=287)
Plasma cfDNA from 129 lung cancer and 158 healthy
*(*Mathios et al., *Nature Communications*, 2021)

Analyzed sequencing
replicates on NovaSeq
platform for technical
variation (n=281)

AHN/DECAMP Lung Cohort (n=57)
Plasma cfDNA from 42 lung cancer, 15 high risk non-cancer individuals
(Bruhm et al., *Nature Genetics*, 2023)

JHU Validation Cohort (n=431)
Plasma cfDNA from 46 lung cancer, 385 healthy
*(*Mathios et al., *Nature Communications*, 2021)

Lung Cancer Monitoring Cohort (n=19)
Plasma cfDNA from 19 lung cancer patients at 75 treatment timepoints
(Phallen et al., *Cancer Research*, 2019; Cristiano et al., *Nature*, 2019)

Lung Cancer Recurrence Cohort (n=25)
Plasma cfDNA from 25 held-out patients in the LUCAS cohort with a prior
history of cancer
*(*Mathios et al., *Nature Communications*, 2021)

Liver Cancer Cohort (n=208)
Plasma cfDNA from 75 liver cancer and 133 high-
risk
(Foda et al., *Cancer Discovery*, 2022)

Multi-cancer Cohort (n=423)
Plasma cfDNA from 215 patients without cancer
and 208 patients with lung* (n=12), bile duct
(n=25), ovarian (n=28), colorectal (n=26), gastric
(n=27), duodenal (n=1), pancreatic (n=34) and
breast (n=54) tumors
*for tissue of origin analyses, supplemented with 19 baseline
timepoints from the Lung Cancer Monitoring Cohort*
*(*Cristiano et al., *Nature*, 2019)

Subset for further analysis

External Validation of Fixed Model

Cross-Validated Model