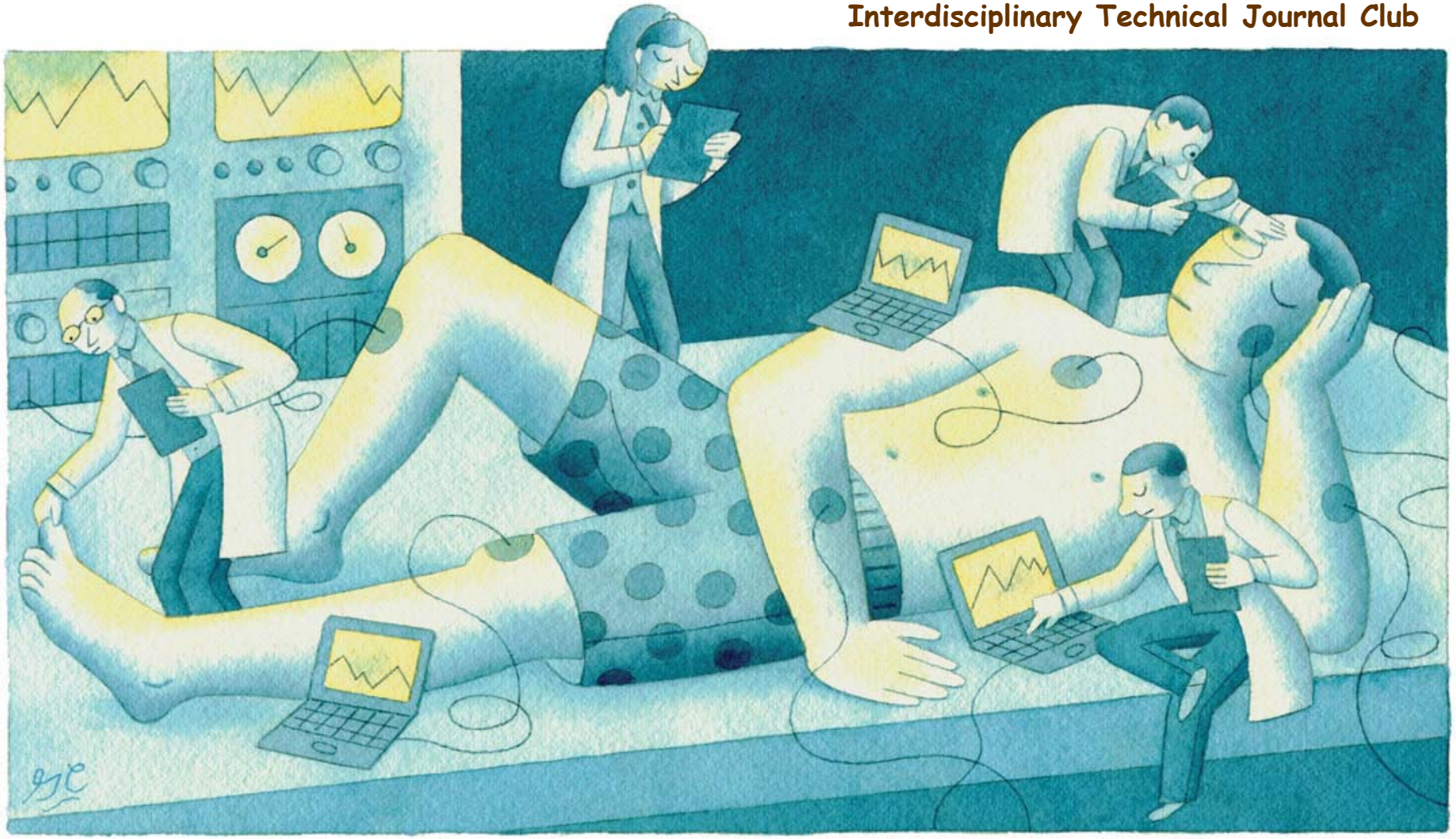
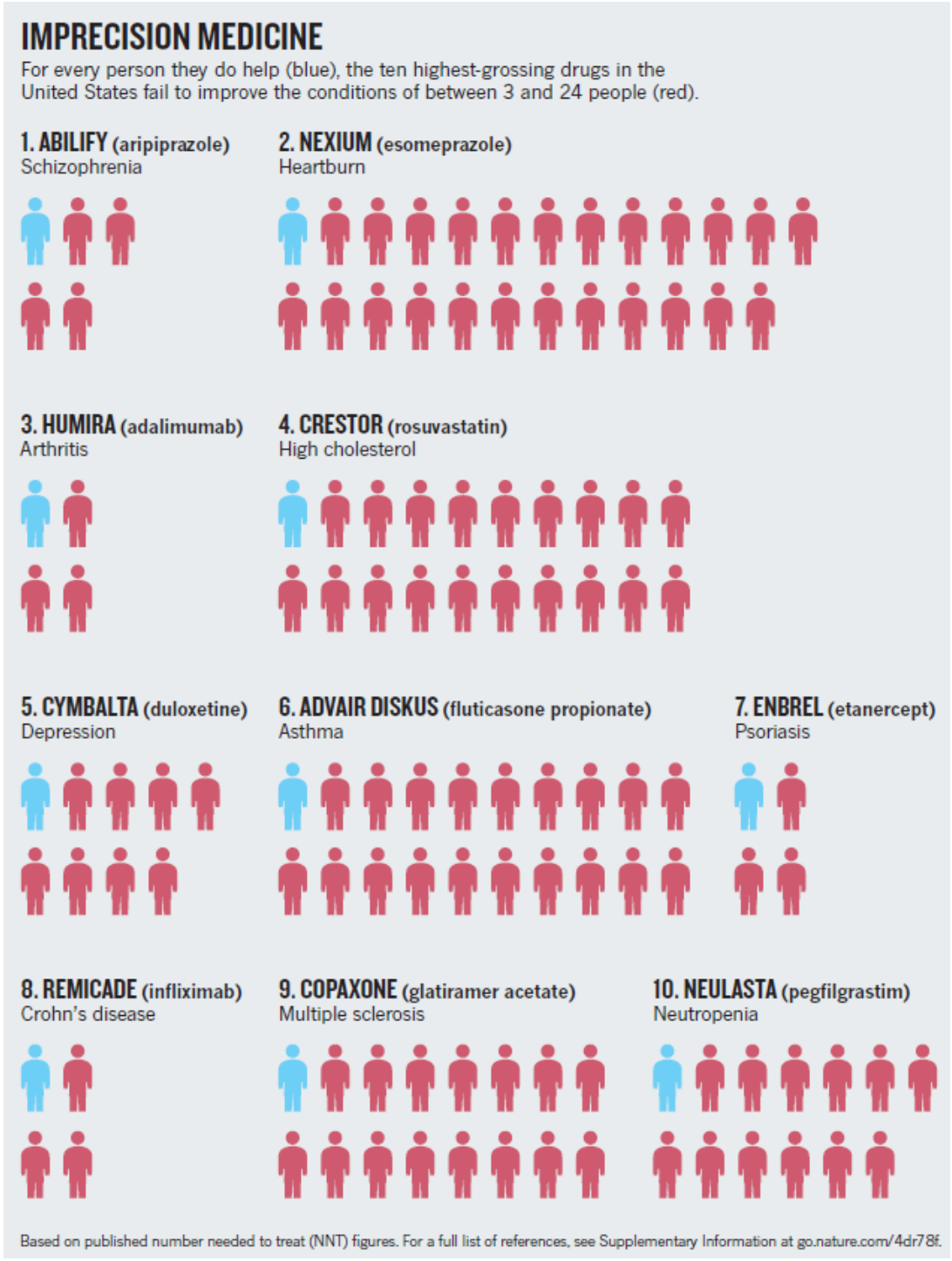
Biomarker development in the era of precision medicine

Bei Li, 23.08.2016

Interdisciplinary Technical Journal Club





The top ten highest-grossing drugs in the United States help between 1 in 25 and 1 in 4 of the people who take them.

- **Blue: help**
- ✘ **Red: not help**

There are even drugs that are harmful to certain ethnic groups.

e.g. an increased mortality associated with regular use of salmeterol were observed among African Americans.

Actually, every day millions of people are taking medications that will not help them.

Recognition that physicians need to take individual variability into account is driving huge interest in **'precision' medicine**.

Nature. 2015 Apr 30;520(7549):609-11.

A brief history of the term ‘precision medicine’

Precision medicine has a long list of predecessor terms with similar meaning:

- Personalized medicine
- P4 (predictive, preventive, participatory and personalized) medicine
- Genomics medicine
- Predictive medicine
- Individualized medicine

Goal of precision medicine

To use molecular data in addition to more traditional clinical information (for example, symptoms, personal history and histology) to tailor medical care to provide the most benefit while minimizing risk.

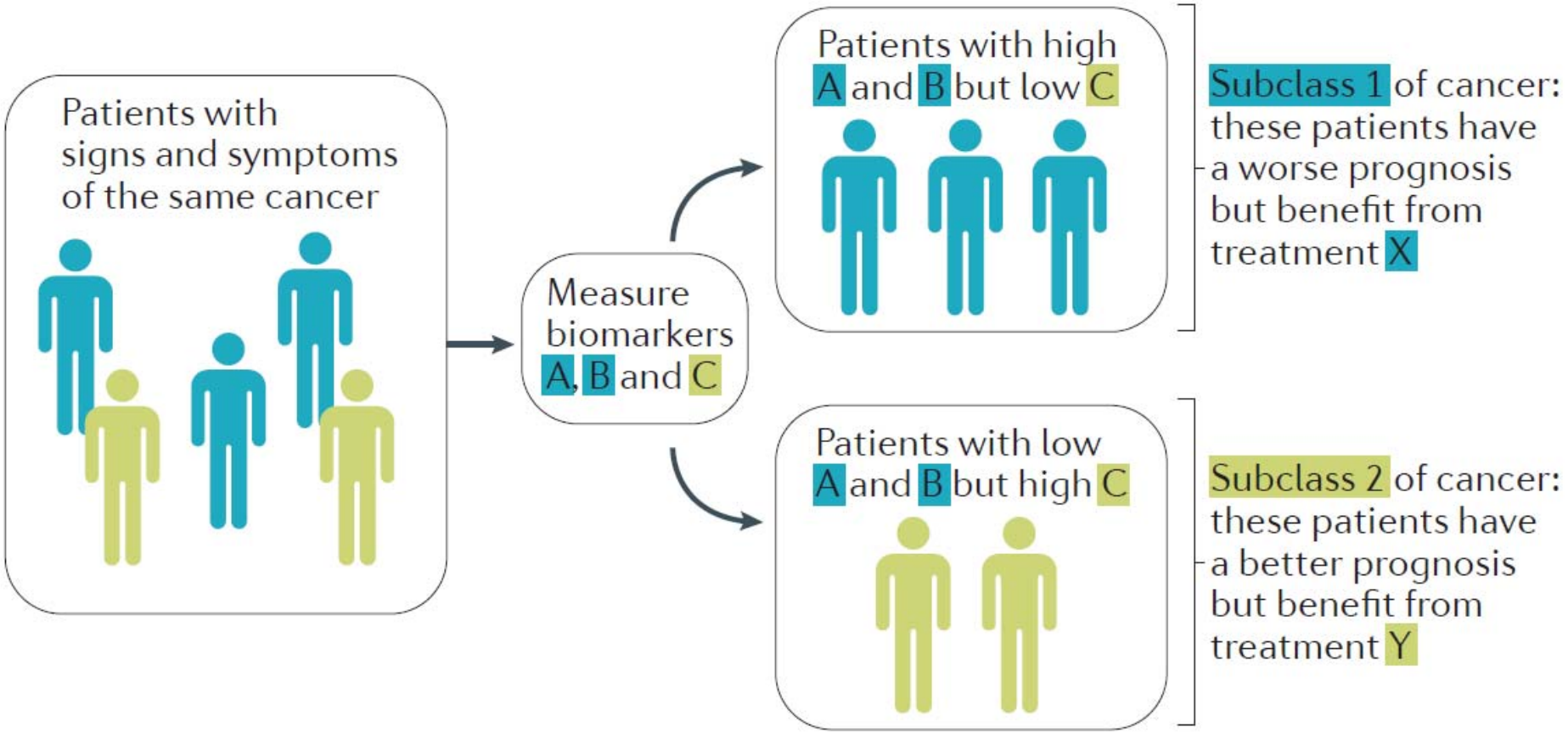
A brief history of the term ‘precision medicine’

Application of precision medicine

- The greatest advances of precision medicine have been achieved in the prediction of response to a drug therapy using companion diagnostics.

- Formalizing and scaling up the precision medicine approach means solving various practical problems, including
  - exploiting the diversity of health-monitoring devices.
  - developing and identifying **new appropriate biomarkers** that can be used to classify patients with the same disease into finer taxa and thus can predict response to a specific drug treatment.

For example, classifying patients into new, specific taxa by biomarkers:



Nat Rev Cancer. 2016 Aug;16(8):525-37.

Different types of molecule as biomarkers

- **Genomic biomarkers**
  - measurement of DNA mutations and translocations, by next-generation DNA sequencing technologies.
- **Transcriptomic biomarkers**
  - tissue specific measurement of mRNA expression, by mRNA sequencing technologies.
- **Epigenomic biomarkers**
  - microRNA (miRNA) and long non-coding RNA (lncRNA), by mRNA sequencing.
  - DNA methylation and histone modifications require immunoprecipitation of the epigenomic mark of interest.
  - DNA methylation uses bisulfite conversion or restriction enzyme before microarray or sequencing analyses.
- **Proteomic biomarkers**
  - immunohistochemical staining of proteins in formalin-fixed, paraffin-embedded tissue samples.
  - tissue microarrays plus immunohistochemical protein stains in a high-throughput manner.
  - mass spectrophotometry.

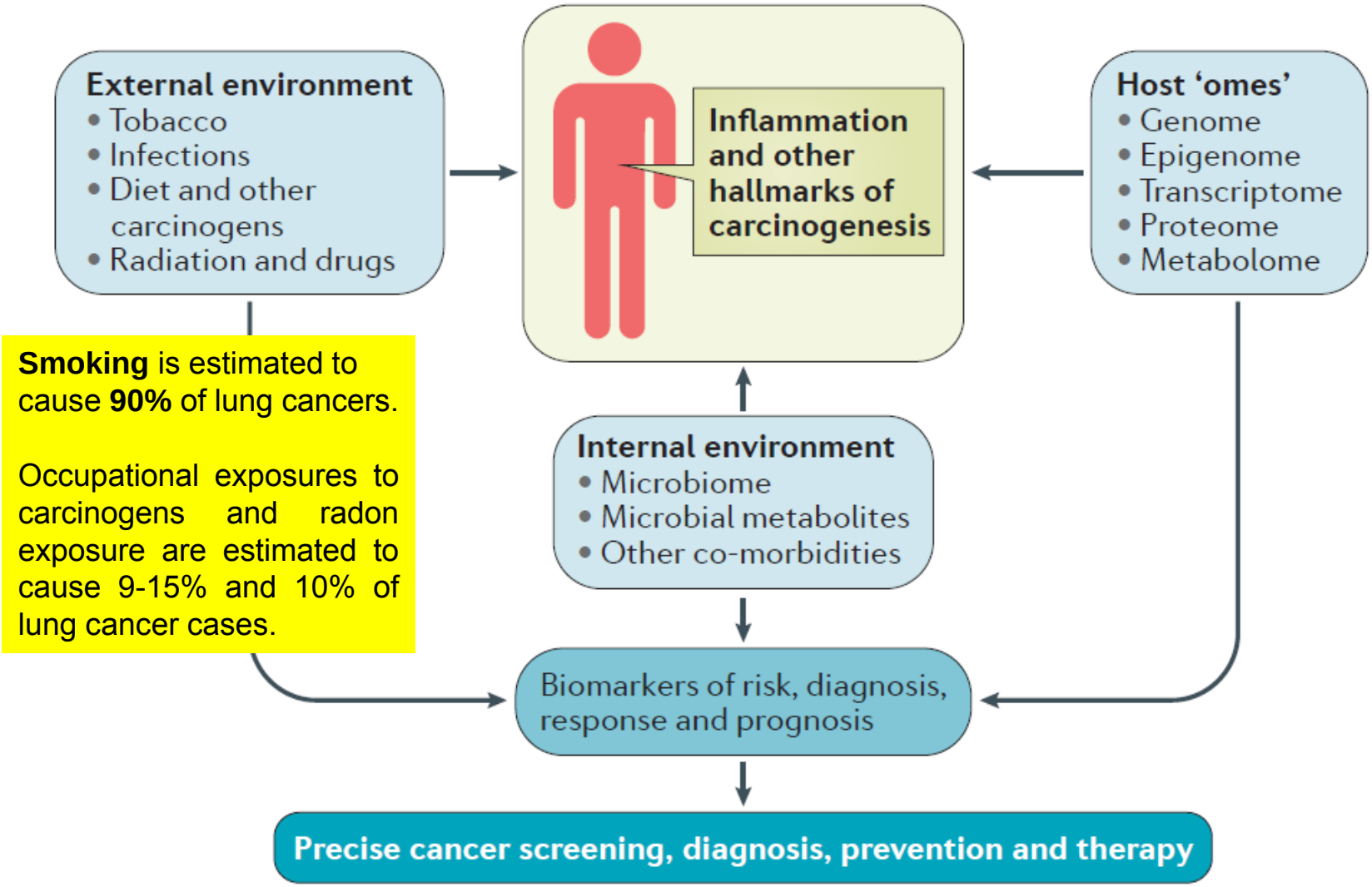
Different types of molecule as biomarkers

- **Metabolomic (also known as metabonomics) biomarkers**
  - use mass spectrophotometry to identify chromatogram peaks as specific metabolites.
  - particularly promising for biomarker development because altered metabolism is considered a hallmark of cancer.
  - metabolomics analyses of blood, urine or faeces could serve as non-invasive biomarkers.

- **Microbiomic biomarkers**
  - culture independent microbial DNA sequencing techniques

  (e.g. cigarette tobacco contains bacteria and that cigarette smoke can disrupt the respiratory tract mucosal barrier to allow microbiota migration into the lung.)

- **Exposome**
  - refers to all types of molecules and events from the environment to which humans can be exposed; for example, drugs, diet or the microbiome.
  - aspects of the exposome are commonly measured by questionnaires, which are administered to patients in the clinic.



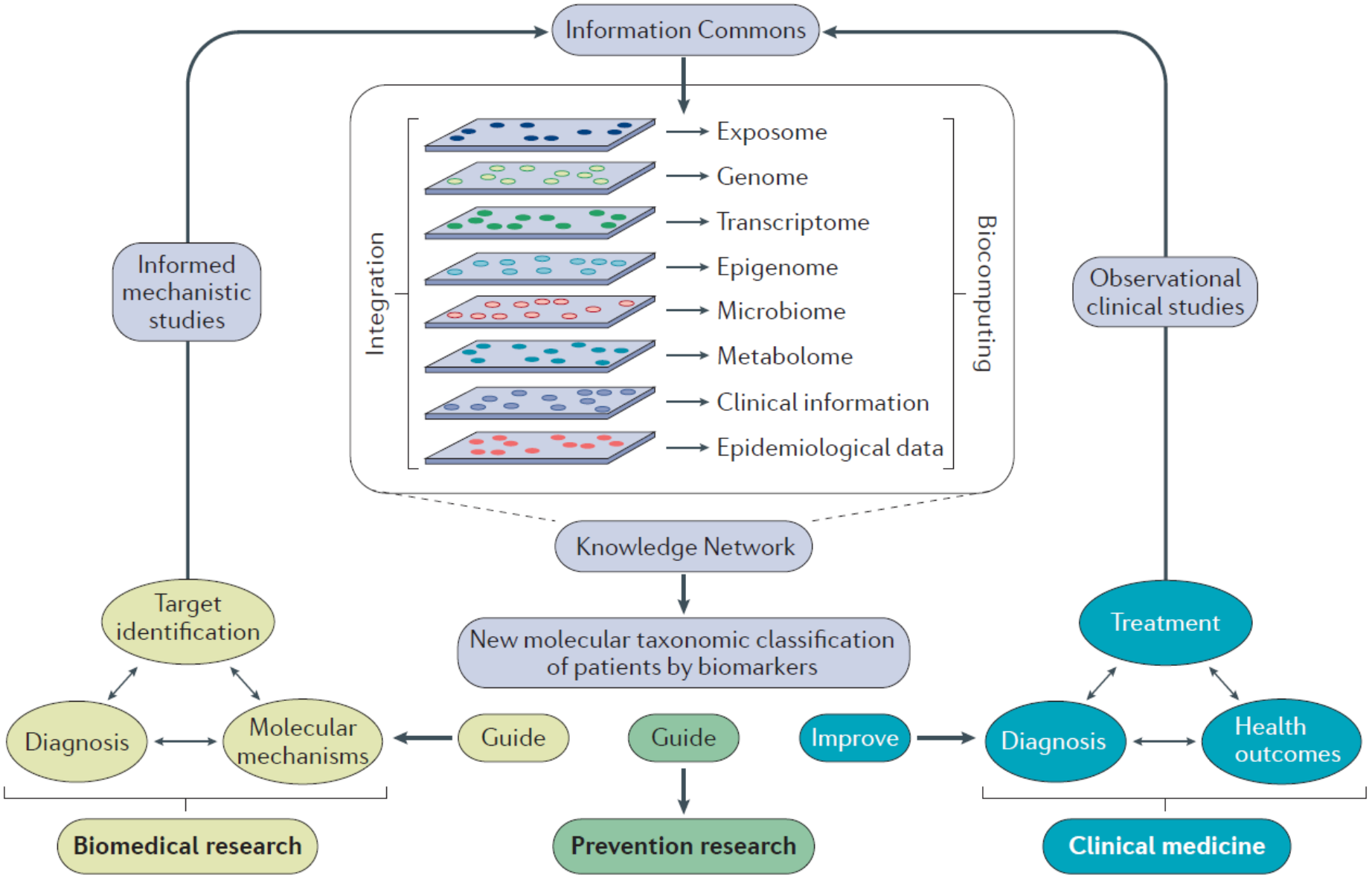
Smoking is estimated to cause **90%** of lung cancers.

Occupational exposures to carcinogens and radon exposure are estimated to cause 9-15% and 10% of lung cancer cases.

E.g. lung exposome comprises a diverse array of molecules and events that come from the external and internal lung environment, which interact with each other and host 'omes' to alter the lung cell environment and may promote or protect against the development of the hallmarks of lung cancer.

Nat Rev Cancer. 2016 Aug;16(8):525-37.

A precision medicine research strategy



Nat Rev Cancer. 2016 Aug;16(8):525-37.

Figure 1. Trends in Age-adjusted Cancer Death Rates* by Site, Males, US, 1930-2012

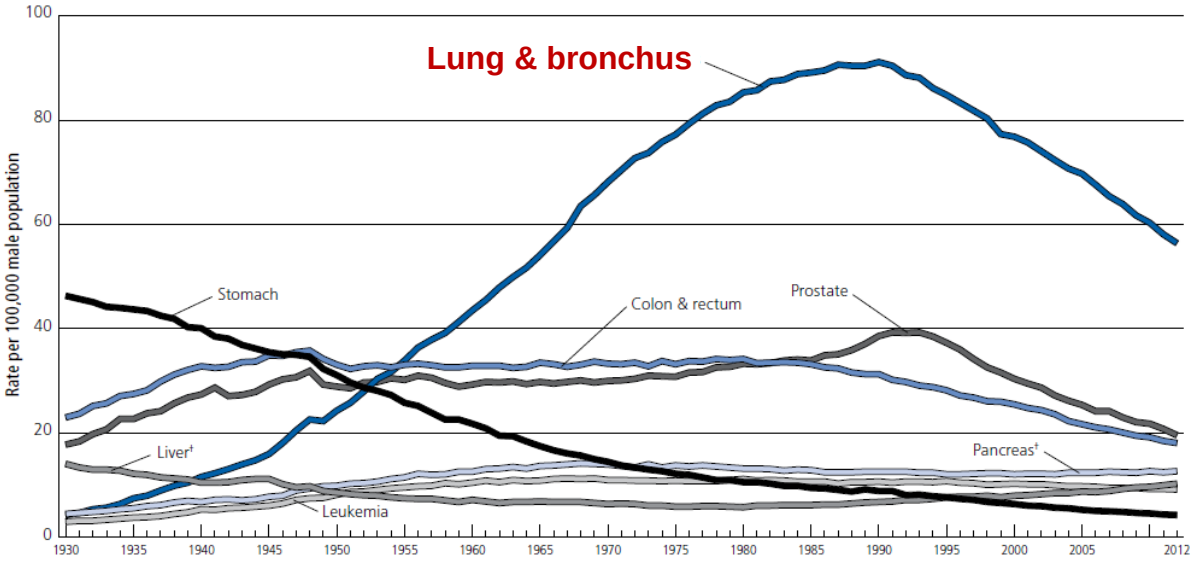
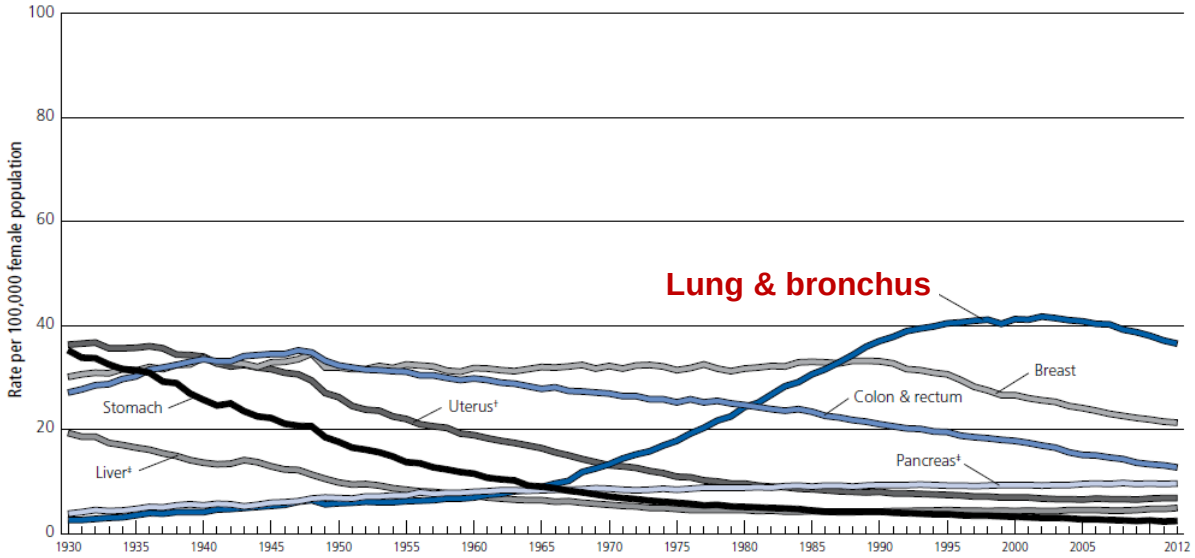


Figure 2. Trends in Age-adjusted Cancer Death Rates* by Site, Females, US, 1930-2012



*Per 100,000, age adjusted to the 2000 US standard population. †Uterus refers to uterine cervix and uterine corpus combined. ‡Mortality rates for pancreatic and liver cancers are increasing.

Note: Due to changes in ICD coding, numerator information has changed over time. Rates for cancers of the liver, lung and bronchus, and colon and rectum are affected by these coding changes.

Source: US Mortality Volumes 1930 to 1959, US Mortality Data 1960 to 2012, National Center for Health Statistics, Centers for Disease Control and Prevention.

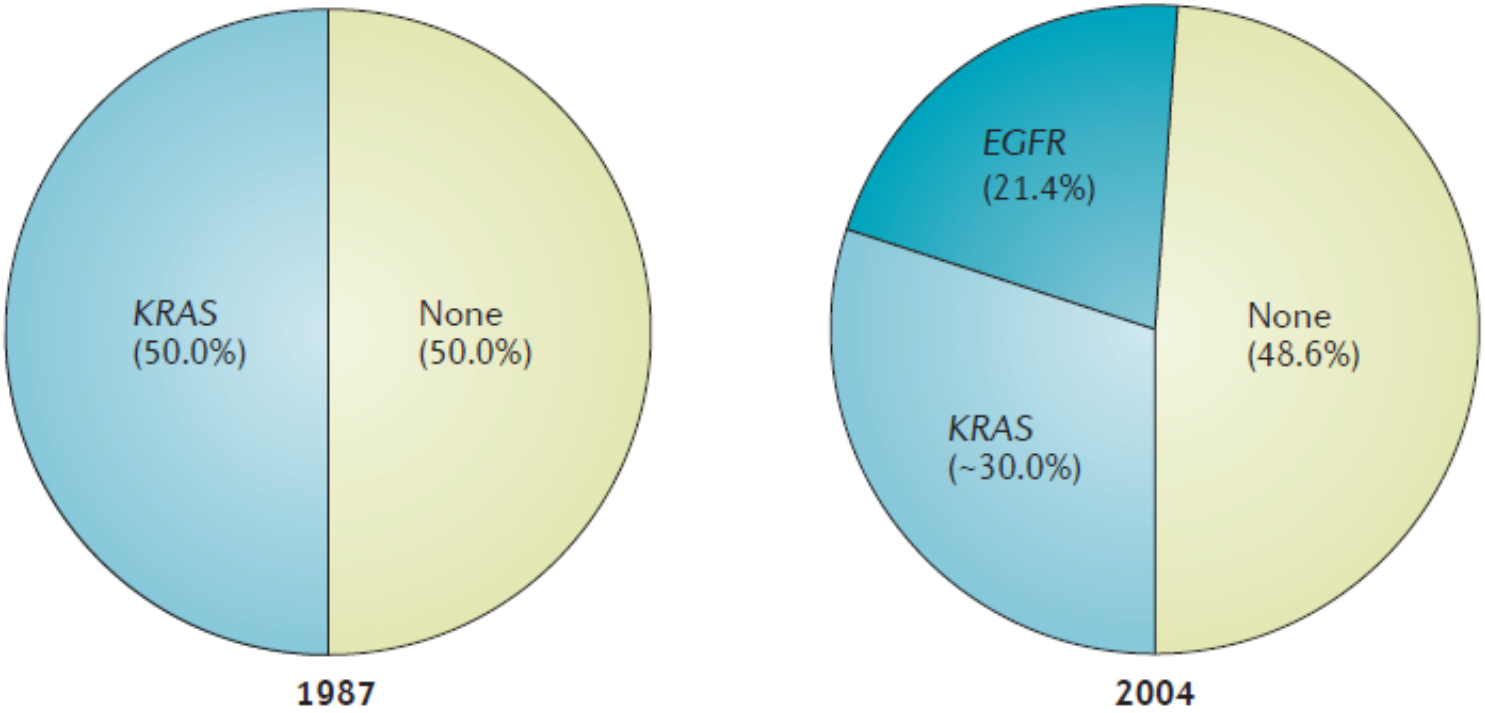
Take lung cancer as an example:

For the year of 2016,

- estimated number of new cases is **224,390**
- estimated number of deaths is **158,080**

Lung cancer causes more deaths worldwide than the other top three cancers combined (**Colon and rectum, Pancreas, Breast**).

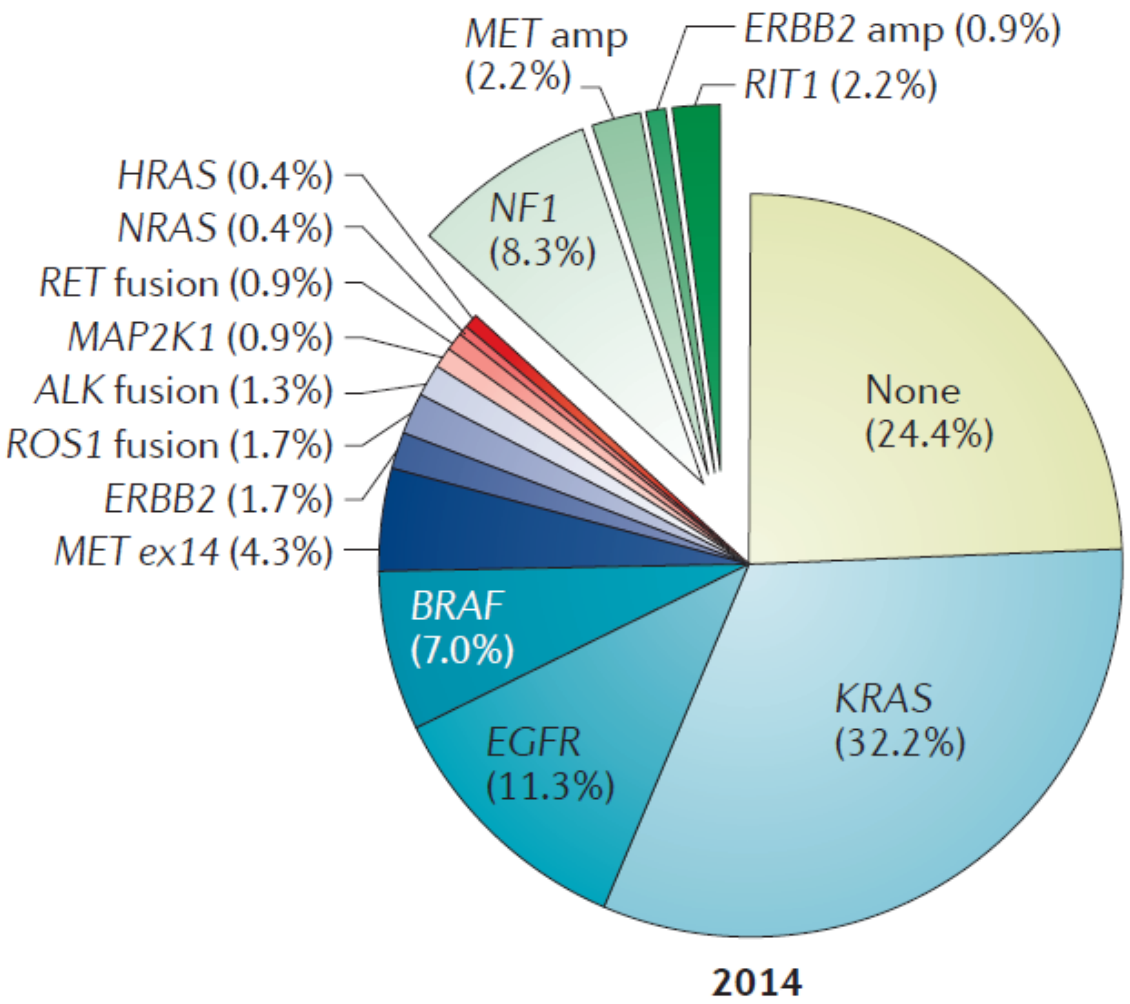
The 5-year survival rate for all lung cancer stages is below 17%, therefore presents have a great need for improved diagnostic precision.



Historically, lung cancer has been grouped into small cell carcinoma, non-small cell squamous cell carcinoma, non-small cell adenocarcinoma and large cell carcinoma subtypes.

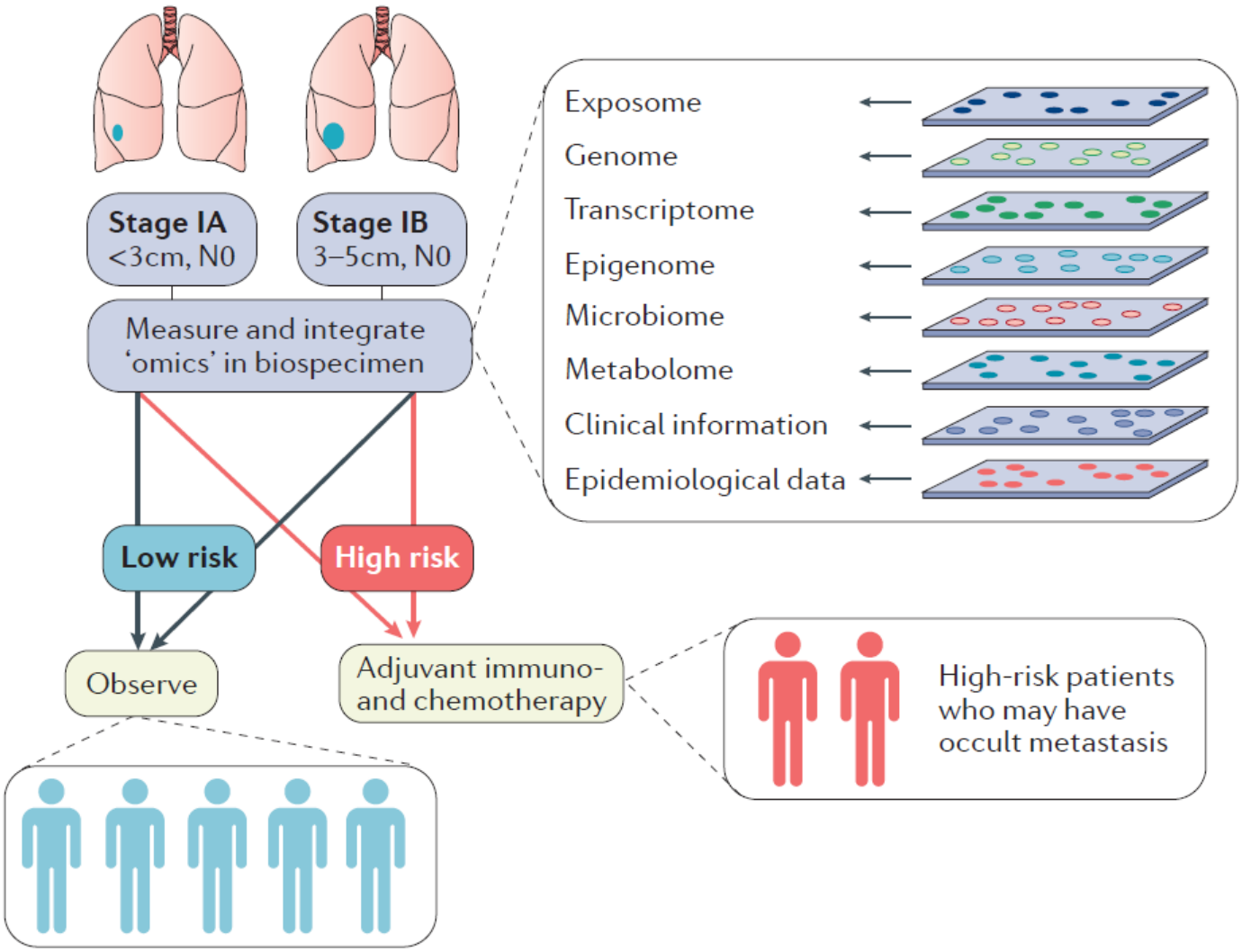
In the late 1980s and the mid-2000s the research community began to recognize that lung adenocarcinoma could further be subdivided beyond histology into cancers that were driven by KRAS and/or EGFR gene mutations.

Nat Rev Cancer. 2016 Aug;16(8):525-37.



In 2014, The Cancer Genome Atlas (TCGA) Network's next-generation sequencing of lung adenocarcinoma led to the identification of more than 15 different gene events that could be exploited for treatment and/or used for subclassifying patients into new taxa.

Nat Rev Cancer. 2016 Aug;16(8):525-37.



E.g. use of precision medicine to classify early stage (IA and IB) lung cancer by biomarkers that predicts risk of recurrence generated into 'low risk for recurrence' and 'high risk for recurrence'. Low-risk patients can be observed post-curative surgery whereas high-risk patients can be provided options for adjuvant therapy post-surgery.

Nat Rev Cancer. 2016 Aug;16(8):525-37.

Challenges of developing a whole country as a national precision medicine cohort

1) Collecting, handling, storing and transporting millions of biospecimens and then analysing these data using multiple different molecular measurement techniques.

2) Collecting electronic medical record data, merging data from different types of medical records and questionnaires, and then storing large amounts of these data.

3) Analysing data from different sources while respecting the strengths and limitations of each type of data.

4) Combining expertise from multiple different disciplines, including clinicians, laboratory researchers, bioinformaticians, biostatisticians and lawyers.

5) Dissemination of these data for researchers to use while ensuring that legal, ethical and privacy concerns of all participants are addressed.

A typical example: Paper-1

ARTICLE

OPEN

doi:10.1038/nature13385

Comprehensive molecular profiling of lung adenocarcinoma

Nature 511, 543–550 (2014).

The Cancer Genome Atlas Research Network*

The list of authors and affiliations contains 53 units:

[1]University of California San Francisco, San Francisco, California 94158, USA. [2]The Eli and Edythe L. Broad Institute, Cambridge, Massachusetts 02142, USA. [3]Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA. [4]Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [5]University of Michigan, Ann Arbor, Michigan 48109, USA. [6]Johns Hopkins University, Baltimore, Maryland 21287, USA. [7]Baylor College of Medicine, Houston, Texas 77030, USA. [8]Washington University, St. Louis, Missouri 63108, USA. [9]Harvard Medical School, Boston, Massachusetts 02115, USA. [10]Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [11]University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [12]University of Texas MD Anderson Cancer Center, Houston, Texas 77054, USA. [13]Princess Margaret Cancer Centre, Toronto, Ontario M5G 2M9, Canada. [14]Brigham and Women's Hospital Boston, Massachusetts 02115, USA. [15]BC Cancer Agency, Vancouver, British Columbia V5Z 4S6, Canada. [16]Mayo Clinic, Rochester, Minnesota 55905, USA. [17]University of Southern California, Los Angeles, California 90033, USA. [18]University of California Santa Cruz, Santa Cruz, California 95064, USA. [19]Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. [20]University of Kentucky, Lexington, Kentucky 40515, USA. [21]Buck Institute for Age Research, Novato, California 94945, USA. [22]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, California 95064, USA. [23]Oregon Health and Science University, Portland, Oregon 97239, USA. [24]International Genomics Consortium, Phoenix, Arizona 85004, USA. [25]Analytical Biological Services, Inc., Wilmington, Delaware 19801, USA. [26]University of Alabama at Birmingham, Birmingham, Alabama 35294, USA. [27]Cleveland Clinic, Cleveland, Ohio 44195, USA. [28]Christiana Care, Newark, Delaware 19713, USA. [29]Cureline, Inc., South San Francisco, California 94080, USA. [30]Emory University, Atlanta, Georgia 30322, USA. [31]Fox Chase Cancer Center, Philadelphia, Philadelphia 19111, USA. [32]ILSbio, Chestertown, Maryland 21620, USA. [33]Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [34]Individumed, Silver Spring, Maryland 20910, USA. [35]The Prince Charles Hospital and the University of Queensland Thoracic Research Center, Brisbane, 4032, Australia. [36]Sullivan Nicolaides Pathology & John Flynn Hospital, Tugun 4680, Australia. [37]Lahey Hospital and Medical Center, Burlington, Massachusetts 01805, USA. [38]NYU Langone Medical Center, New York, New York 10016, USA. [39]Ontario Tumour Bank, Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada. [40]Penrose St. Francis Health Services, Colorado Springs, Colorado 80907, USA. [41]Roswell Park Cancer Center, Buffalo, New York 14263, USA. [42]Rush University Medical Center, Chicago, Illinois 60612, USA. [43]St. Petersburg Academic University, St Petersburg 199034, Russia. [44]Thoraxklinik am Universitätsklinikum Heidelberg, 69126 Heidelberg, Germany. [45]University Heidelberg, 69120 Heidelberg, Germany. [46]University of Cologne, 50931 Cologne, Germany. [47]University of Miami, Sylvester Comprehensive Cancer Center, Miami, Florida 33136, USA. [48]University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. [49]Center Hospitalier Universitaire Vaudois, Lausanne and European Thoracic Oncology Platform, CH-1011 Lausanne, Switzerland. [50]Ziauddin University Hospital, Karachi, 75300, Pakistan. [51]SRA International, Inc., Fairfax, Virginia 22033, USA. [52]National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [53]National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

A typical example: Paper-1

ARTICLE

OPEN

doi:10.1038/nature13385

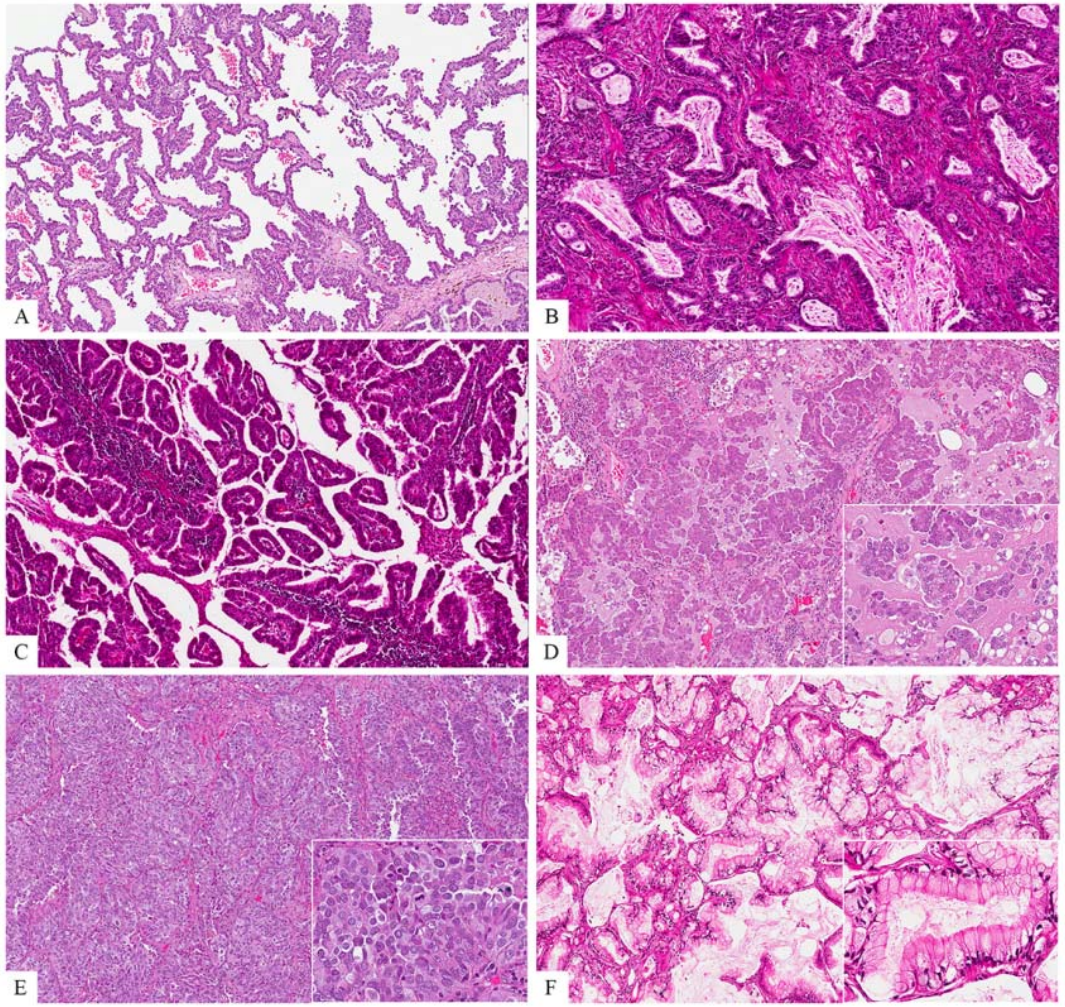
Comprehensive molecular profiling of lung adenocarcinoma

The Cancer Genome Atlas Research Network*

Nature 511, 543–550 (2014).

They reported molecular profiling of 230 resected lung adenocarcinomas using **messenger RNA**, **microRNA** and **DNA sequencing** integrated with **copy number, methylation and proteomic analyses**, with attention towards pathobiology and clinically actionable events.

(1) Clinical samples and histopathologic data

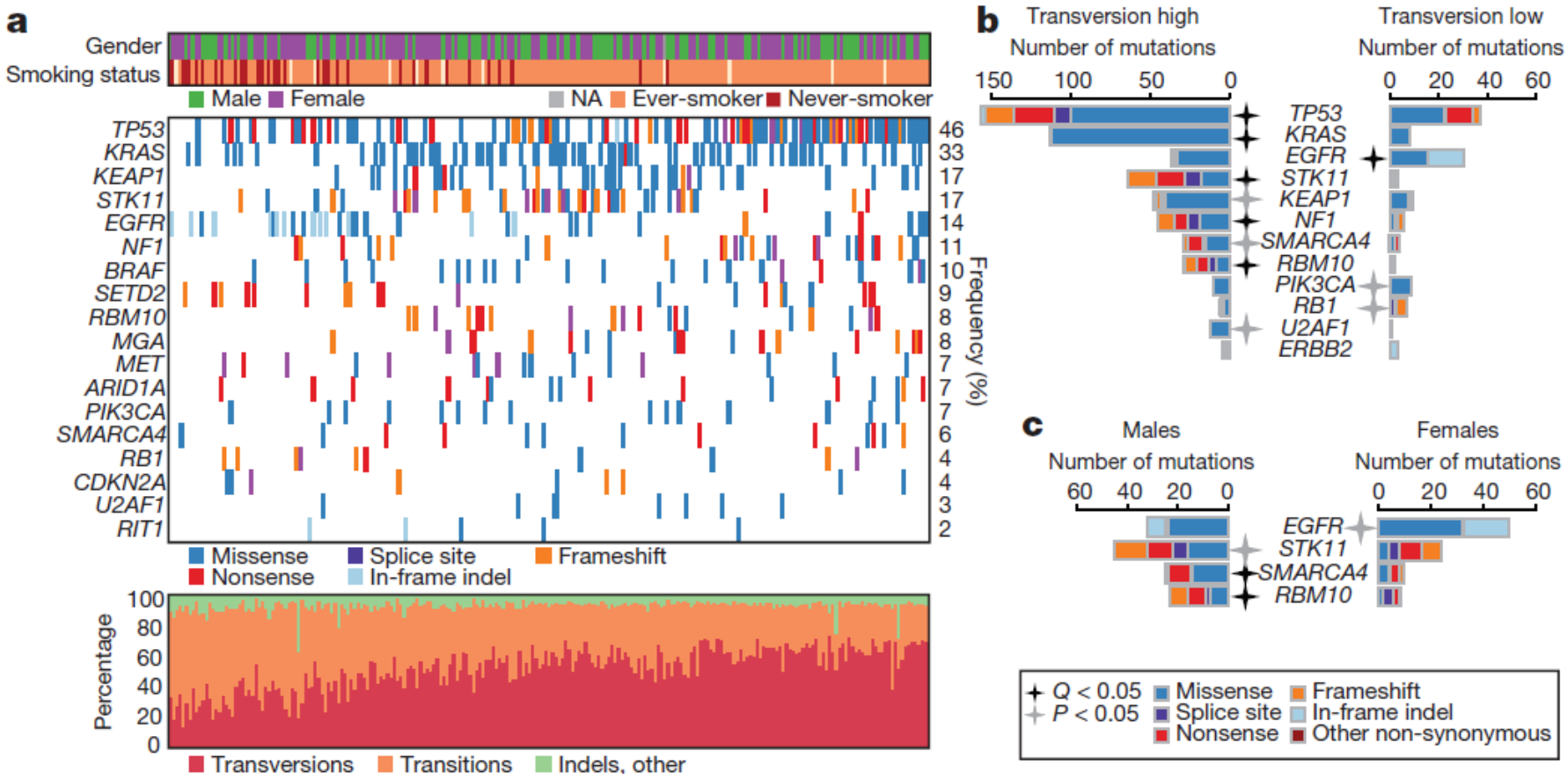


They analysed tumour and matched normal material from **230** previously untreated lung adenocarcinoma patients who provided informed consent.

All major histologic types of lung adenocarcinoma were represented:

A) 5% lepidic
B) 33% acinar
C) 9% papillary
D) 14% micropapillary
E) 25% solid
F) 4% invasivemucinous, (0.4%colloid and 8% unclassifiable adenocarcinoma)

(2) Somatically acquired DNA alterations



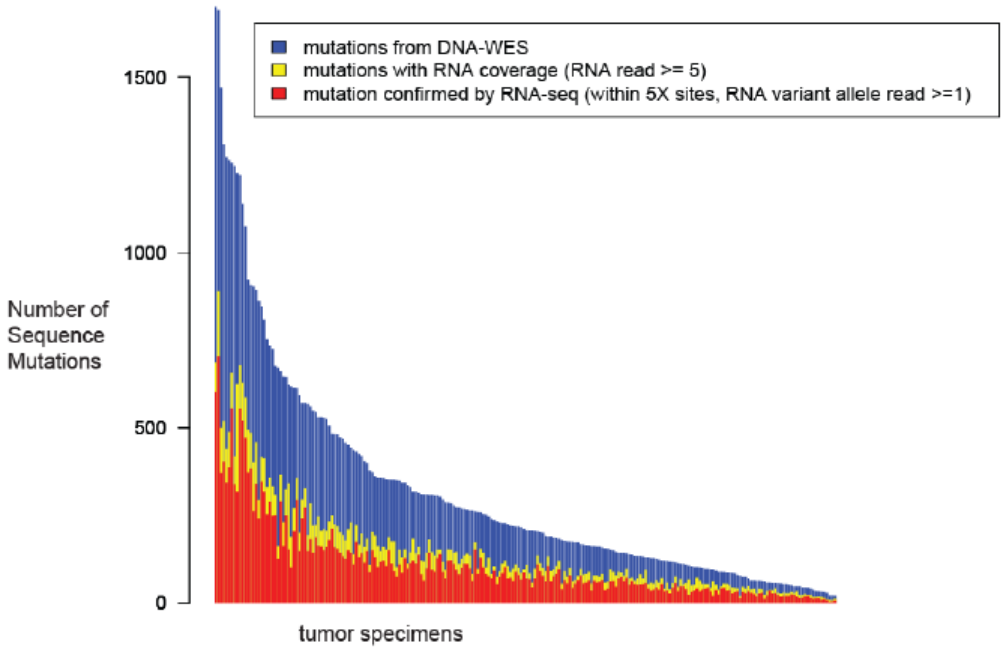
Somatic mutations in lung adenocarcinoma (Whole-exome sequencing (WES) on tumour and germline DNA)

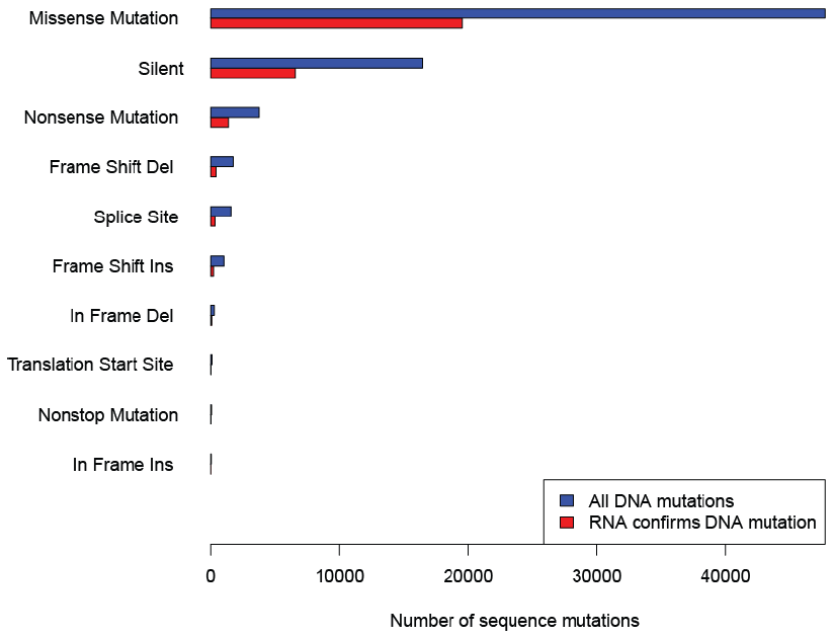
a) Analysis of these tumour/normal pairs highlighted **18 statistically significant mutated genes.**

b) The transversion-high and transversion-low patient cohorts harboured different gene mutations.

c) Only a fraction of significantly mutated genes were enriched in men or women.

(3) Description of aberrant RNA transcripts



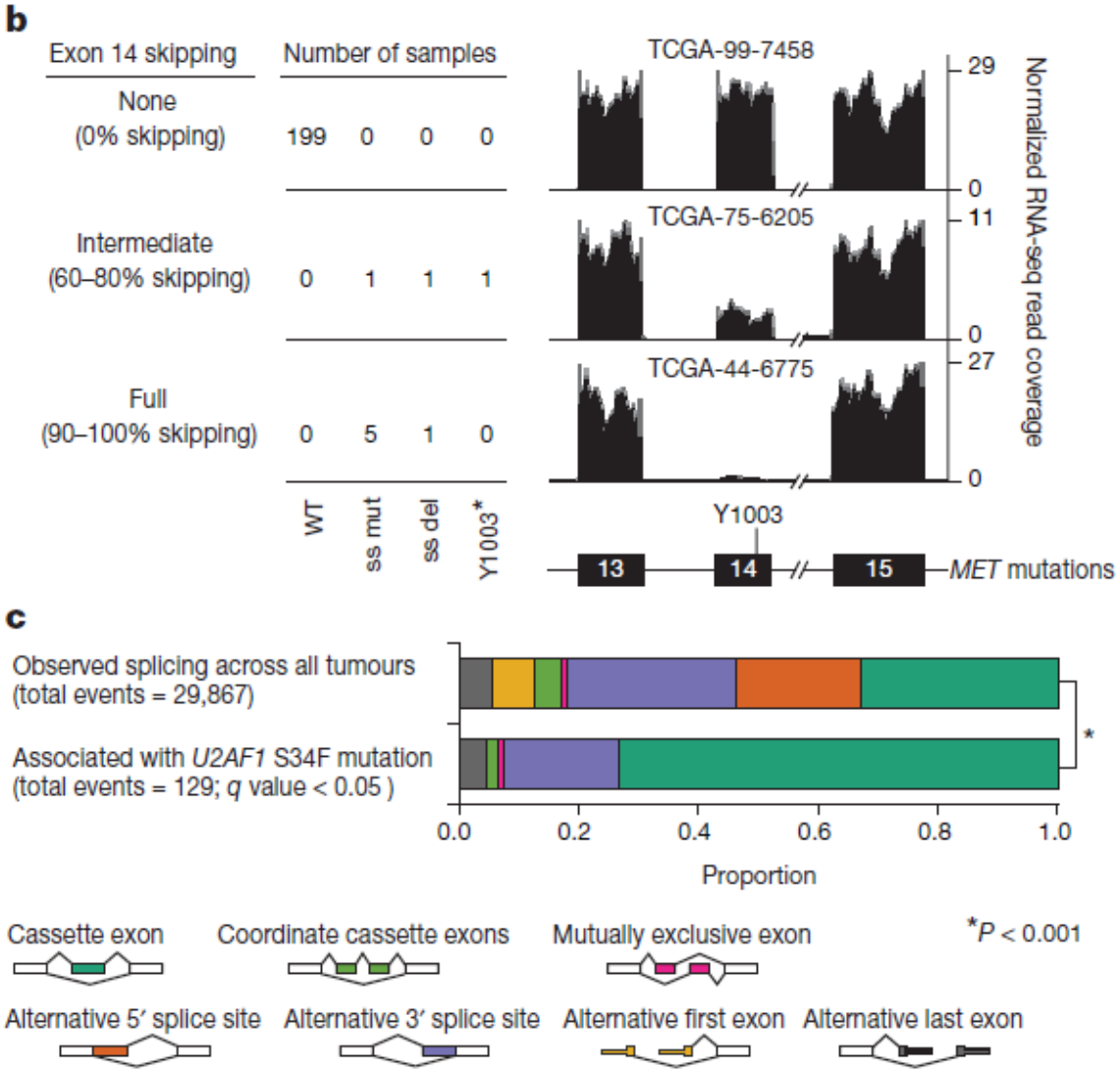


DNA-WES sequence mutation validation by RNA-seq

Combining DNA with mRNA sequencing enabled them to catalogue aberrant RNA transcripts and to identify the DNA-encoded mechanism for the aberration.

75% of somatic mutations identified by WES were present in the RNA transcriptome when the locus in question was expressed.

(3) Description of aberrant RNA transcripts



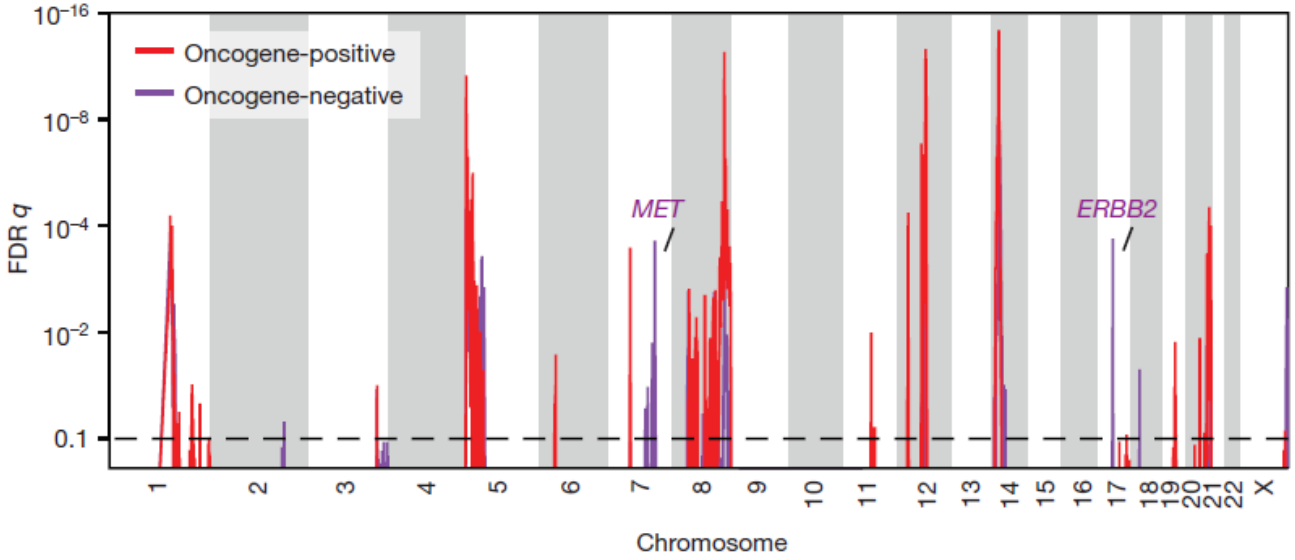
For example, MET exon 14 skipping was observed in the presence of exon 14 splice site mutation (ss mut), splice site deletion (ss del) or a Y1003* mutation.

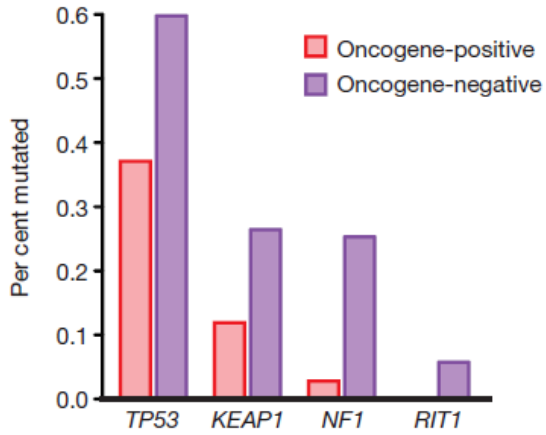
A total of 22 samples had insufficient coverage around exon 14 for quantification.

129 splicing events were identified strongly associated with *U2AF1*^{S34F} mutation.

Cassette exons and alternative 3' splice sites were most commonly affected.

(4) Candidate driver genes



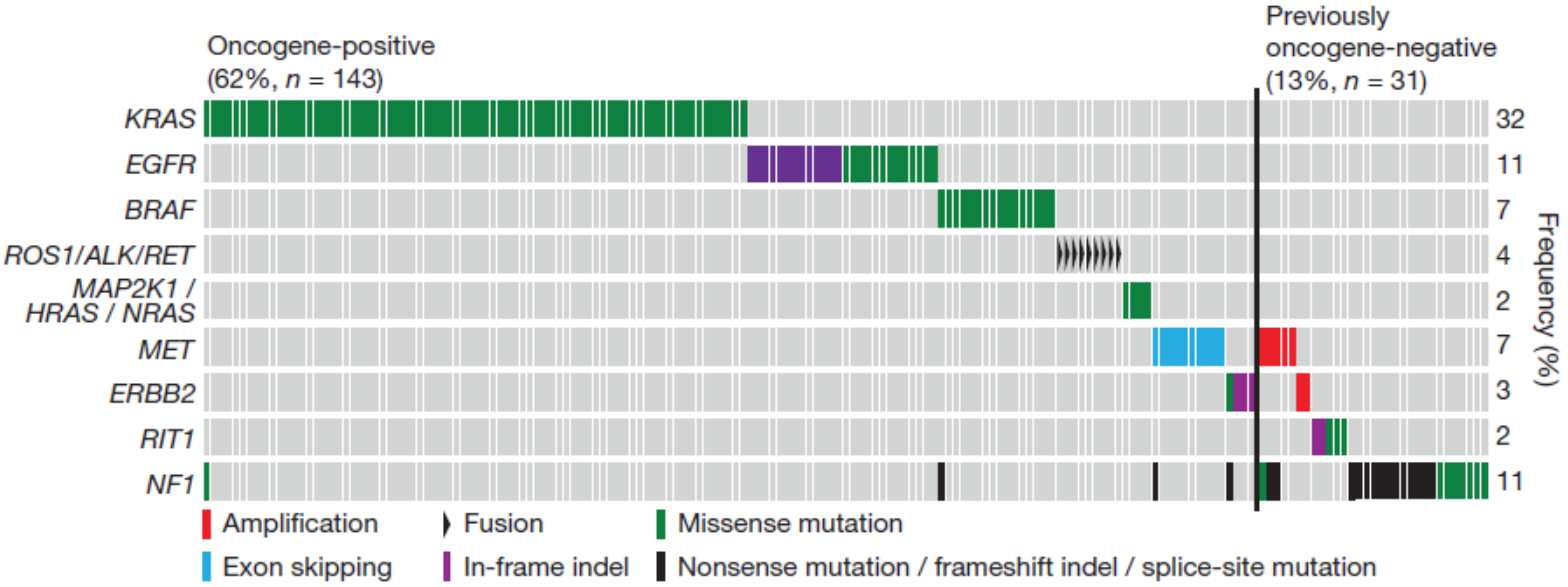


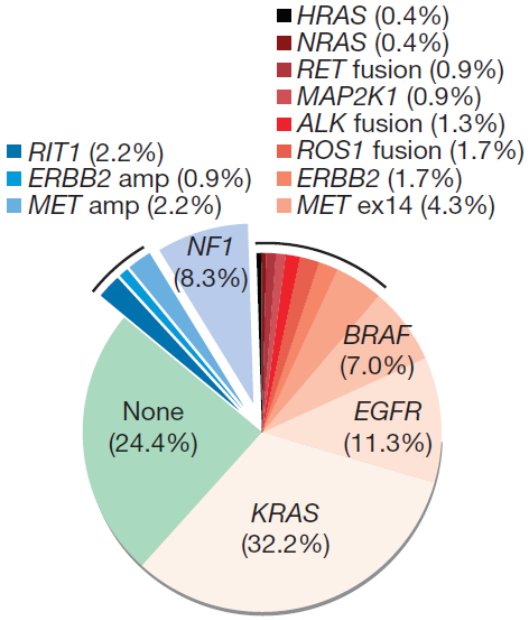
GISTIC analysis of focal amplifications in oncogene-negative (n=87) and oncogene-positive (n=143) samples identified unique focal *ERBB2* and *MET* amplifications in the oncogene-negative subset.

TP53, KEAP1, NF1 and RIT1 mutations were significantly enriched in oncogene-negative tumours.

"GISTIC is a tool to identify genes targeted by somatic copy-number alterations (SCNAs) that drive cancer growth. By separating SCNA profiles into underlying arm-level and focal alterations, GISTIC estimates the background rates for each category as well as defines the boundaries of SCNA regions."

(4) Candidate driver genes



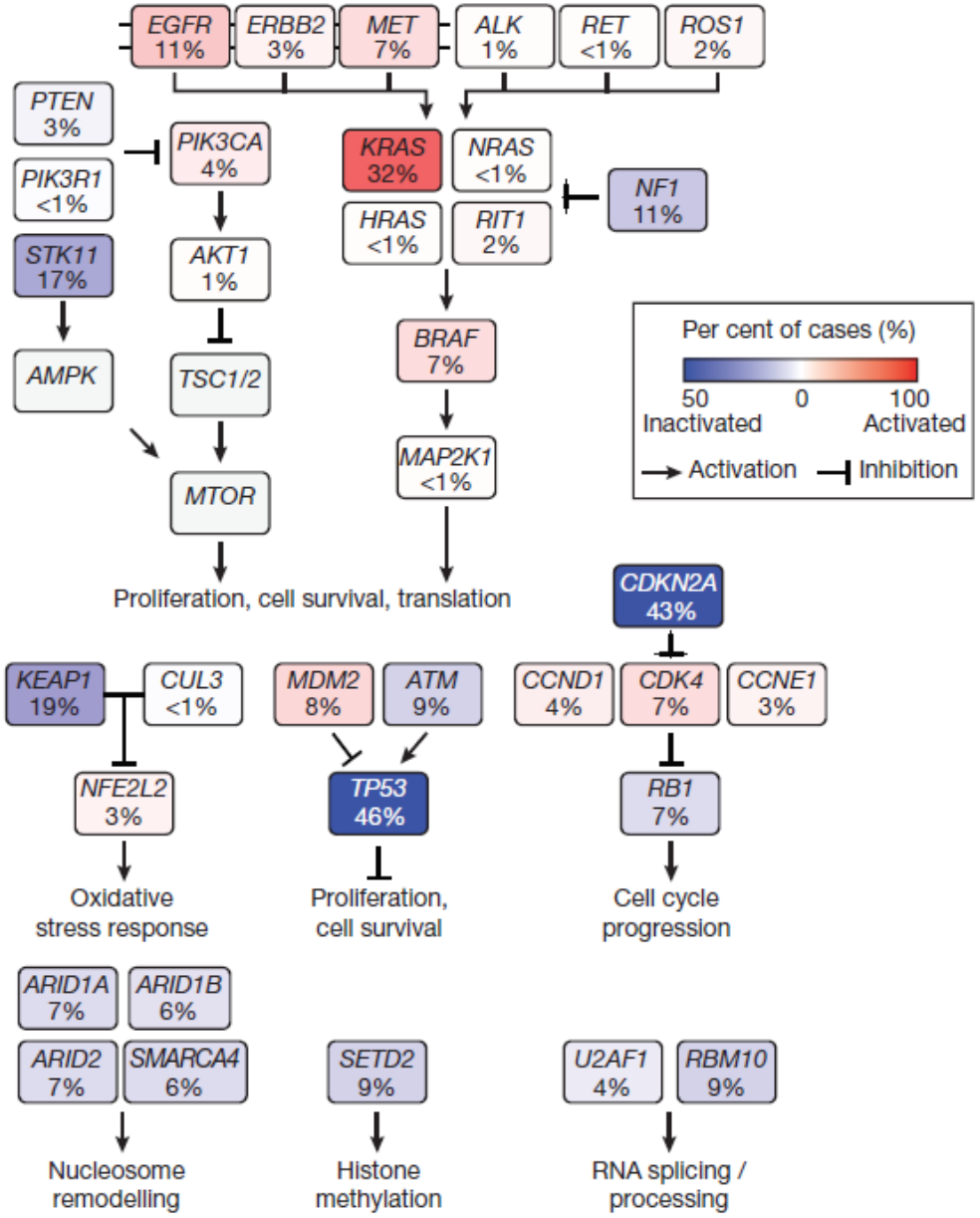


The receptor tyrosine kinase (RTK)/RAS/RAF pathway is frequently mutated in lung adenocarcinoma.

They showed co-mutation plot of variants of known significance within the RTK/RAS/RAF pathway in lung adenocarcinoma.

New candidate driver oncogenes (blue) and known somatically activated drivers events (red) that activate the RTK/RAS/RAF pathway can be found in the majority of the 230 lung adenocarcinomas.

(5) Recurrent alterations in multiple key pathways

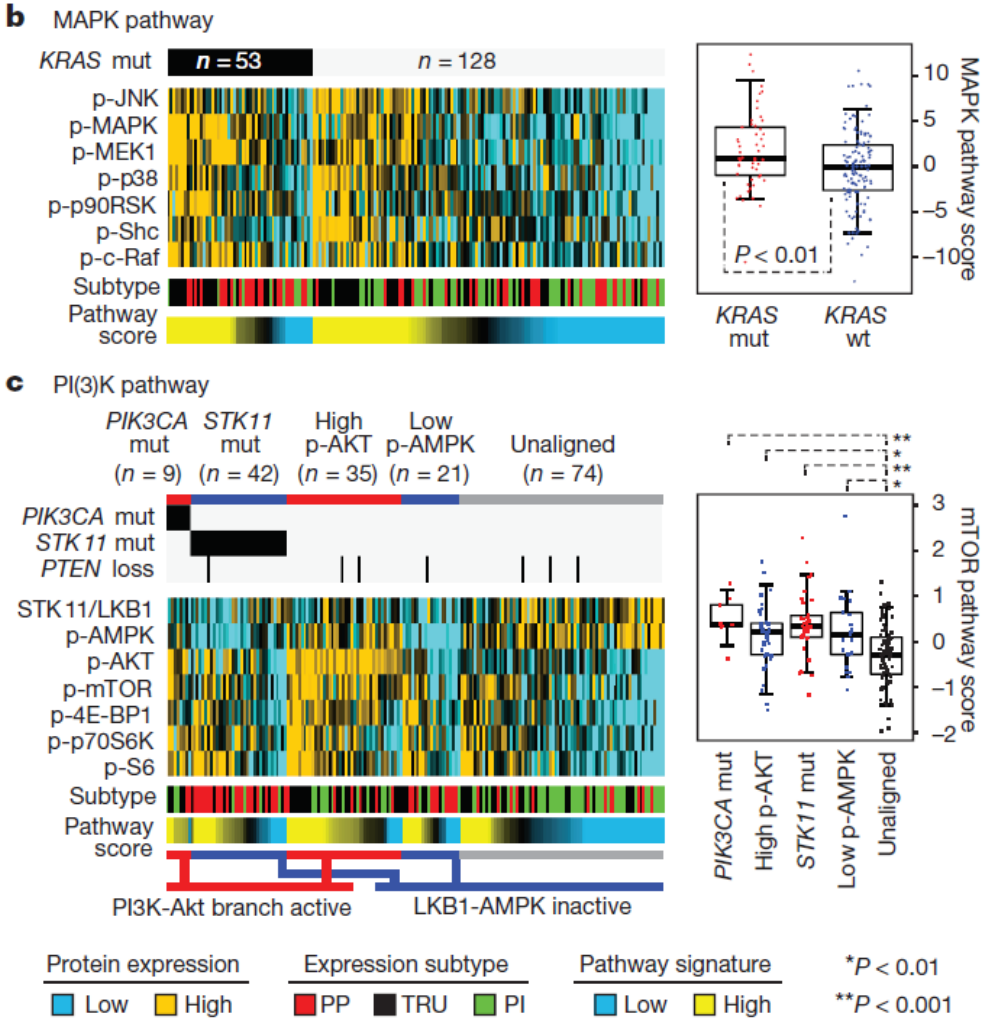


Recurrent aberrations in multiple key pathways and processes were used to characterize lung adenocarcinoma.

The key pathway alterations in lung adenocarcinoma were

- RTK/RAS/RAF signalling (76%)
- PI(3)K-mTOR signalling (25%)
- p53 pathway (63%)
- cell cycle regulators (64%)
- oxidative stress response (22%)
- mutation of various chromatin and RNA splicing factors (49%)

(5) Recurrent alterations in key pathways



Reverse-phase protein arrays (RPPA) provided proteomic and phosphoproteomic phenotypic evidence of pathway activity.

e.g. although KRAS-mutant lung adenocarcinomas had higher levels of phosphorylated MAPK than KRAS wild-type tumours had on average, many KRAS wild-type tumours displayed significant MAPK pathway activation, suggesting additional, still undetected RTK/RAS/ RAF pathway alterations.

This analysis suggested that DNA sequencing did not identify all samples with phosphoprotein evidence of activation of a given signalling pathway.

"RPPA is a protein array designed as a micro- or nano-scaled dot-blot platform that allows measurement of protein expression levels in a large number of biological samples simultaneously in a quantitative manner when high-quality antibodies are available."

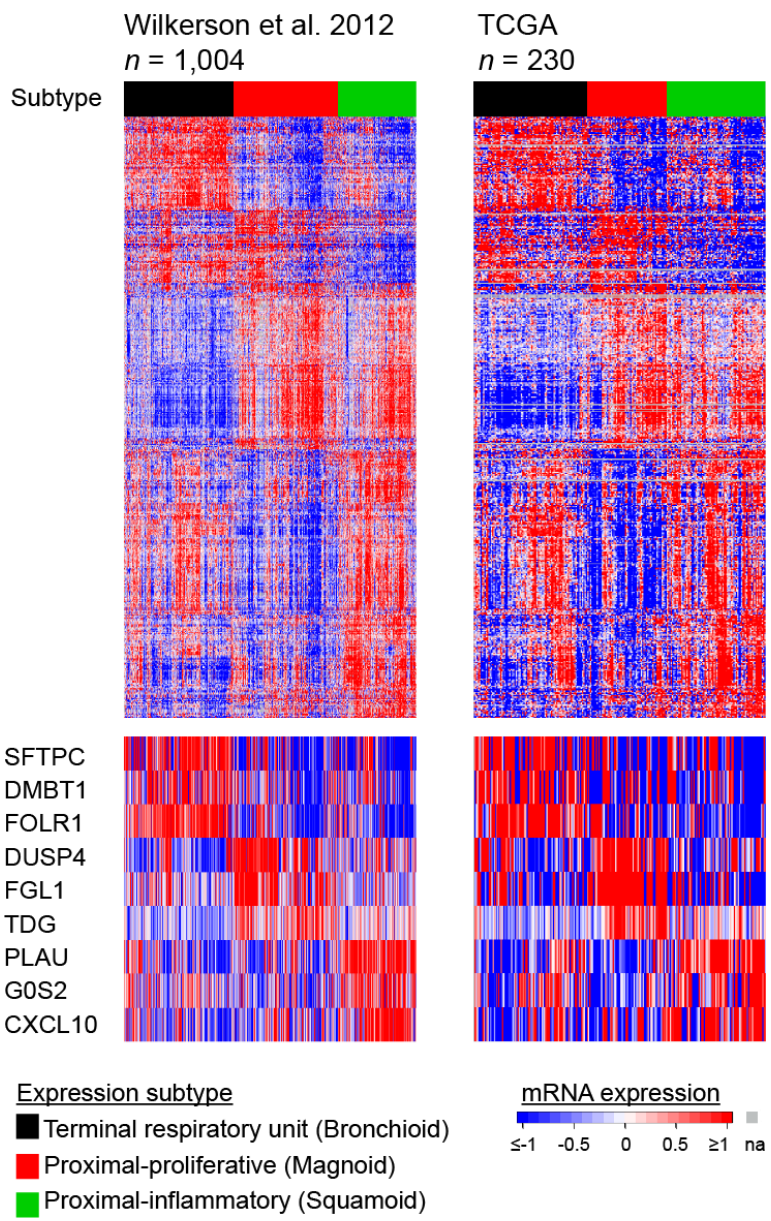
(6) Molecular subtypes of lung adenocarcinoma

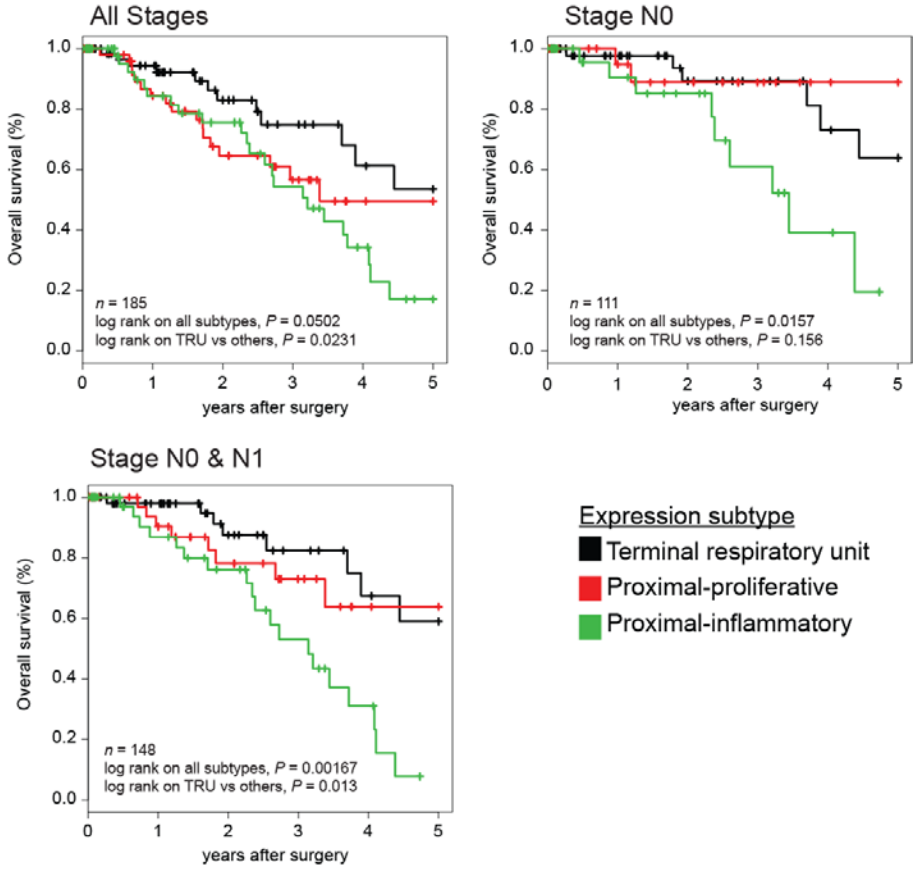
Broad transcriptional and epigenetic profiling can reveal downstream consequences of driver mutations, provide clinically relevant classification and offer insight into tumours lacking clear drivers.

To coordinate naming of the transcriptional subtypes with the histopathological, anatomic and mutational classifications of lung adenocarcinoma, they proposed an updated nomenclature of transcriptional subtypes:

- **the terminal respiratory unit (TRU, formerly bronchioid),**
- **the proximal-inflammatory (PI, formerly squamoid),**
- **the proximal-proliferative (PP, formerly magnoid)**

(6) Molecular subtypes of lung adenocarcinoma

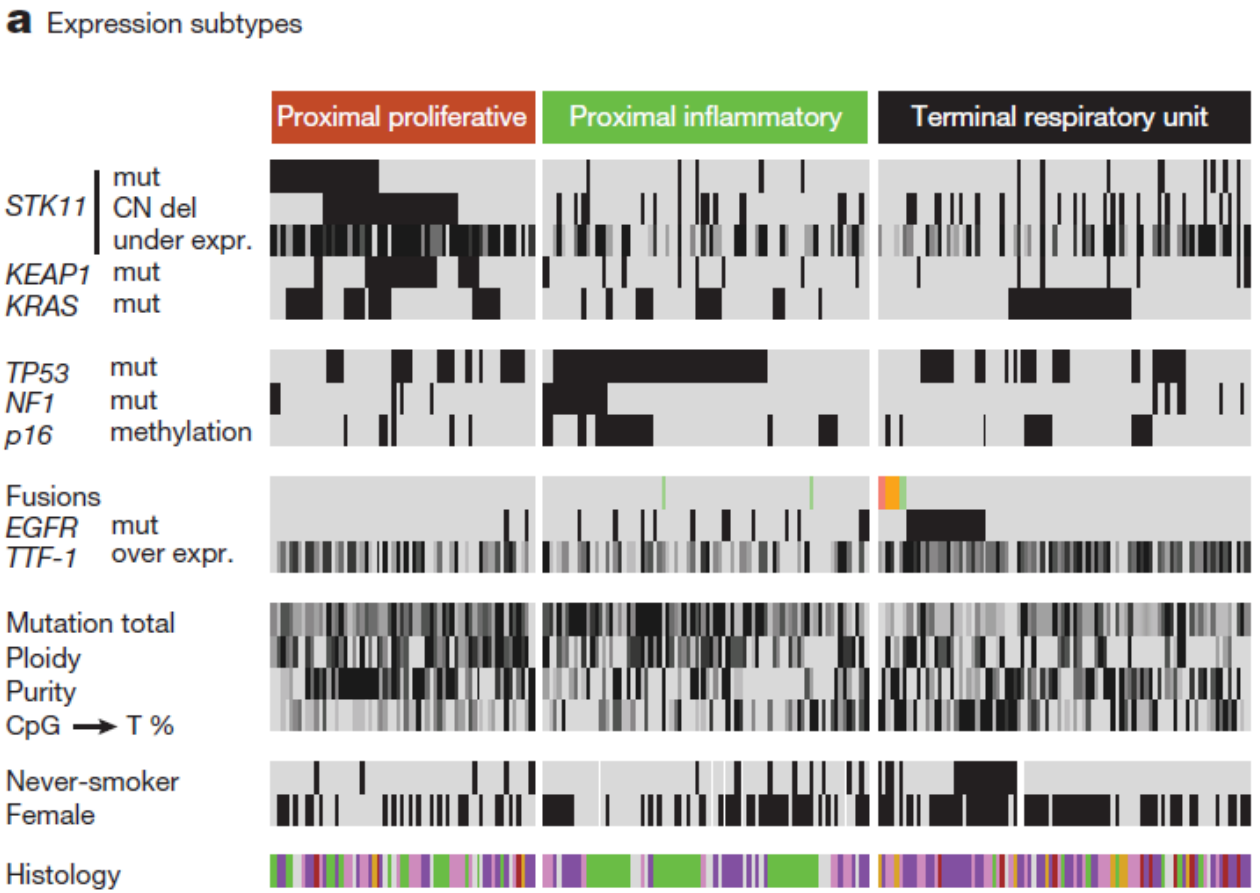




Previously reported associations of expression signatures with pathways and clinical outcomes were observed and integration with multi-analyte data revealed statistically significant genomic alterations associated with these transcriptional subtypes.

TRU subtype membership was prognostically favourable.

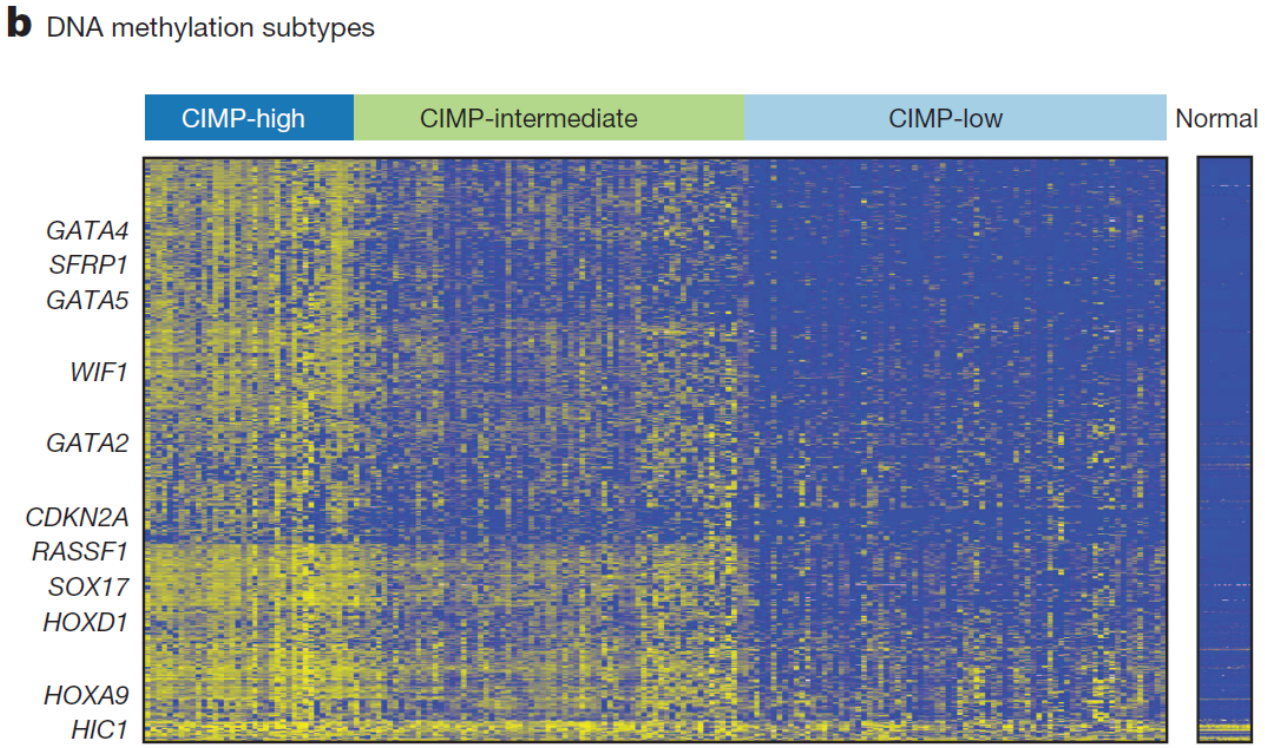
(6) Molecular subtypes of lung adenocarcinoma



(Tumours are displayed as columns, grouped by **mRNA expression subtypes**)

The subtypes exhibited different mutation rates, transition frequencies, genomic ploidy profiles, patterns of large-scale aberration, and differed in their association with smoking history.

(6) Molecular subtypes of lung adenocarcinoma

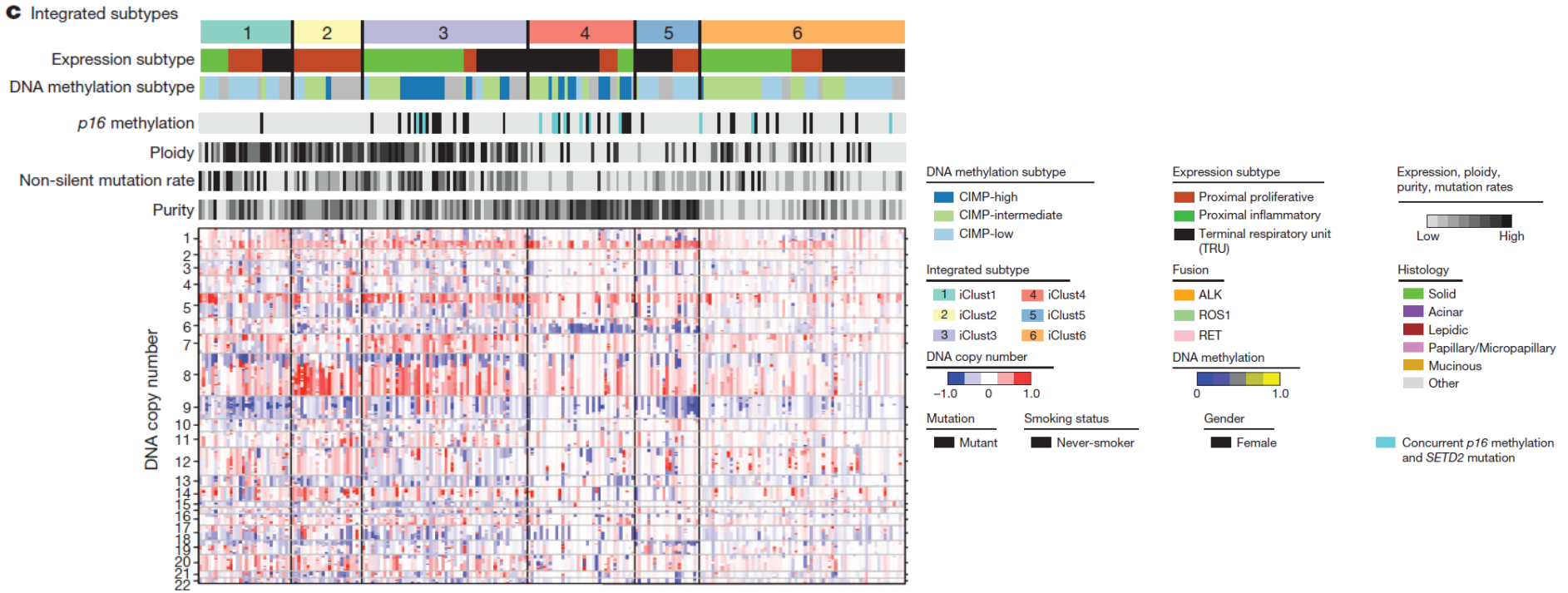


(Tumours are displayed as columns, grouped by **DNA methylation subtypes**)

To examine chromatin states in an unbiased manner, they selected the most variable DNA methylation-specific probes in CpG island promoter regions and clustered them by methylation intensity.

This analysis divided samples into two distinct subsets: a significantly altered CpG island methylator phenotype high (CIMP-High) cluster and a more normal-like (CIMP-Low) group, with a third set of samples occupying an intermediate level of methylation at CIMP sites.

(6) Molecular subtypes of lung adenocarcinoma



(Tumors are displayed as columns, grouped by **integrated subtypes**)

Integrative clustering (by iCluster analysis) of copy number, DNA methylation and mRNA expression data found six clusters.

- Tumour ploidy and mutation rate are higher in clusters 1–3 than in clusters 4–6.
- Clusters 1–3 frequently harbour *TP53* mutations and are enriched for the two proximal transcriptional subtypes.

Summary of the paper

It showed a comprehensive analysis of lung adenocarcinoma describing many different molecular subtypes.

- assessed the mutation profiles, structural rearrangements, copy number alterations, DNA methylation, mRNA, miRNA and protein expression of 230 lung adenocarcinomas.

- implicated both chromatin modifications and splicing alterations in lung adenocarcinoma through the integration of DNA, transcriptome and methylome analysis.

- provide new knowledge by illuminating modes of genomic alteration, highlighting previously unappreciated altered genes, and enabling further refinement in sub-classification for the improved personalization of treatment for this deadly disease.

Another example: Paper-2

ORIGINAL ARTICLE

An Integrated Prognostic Classifier for Stage I Lung Adenocarcinoma Based on mRNA, microRNA, and DNA Methylation Biomarkers

Ana I. Robles, PhD, Eri Arai, MD,† Ewy A. Mathé, PhD*, Hirokazu Okayama, MD, PhD,* Aaron J. Schetter, PhD, MPH,* Derek Brown, MS,* David Petersen, BS,‡ Elise D. Bowman, MS,* Rintaro Noro, MD,* Judith A. Welsh, BS,* Daniel C. Edelman, PhD,‡ Holly S. Stevenson, PhD,‡ Yonghong Wang, PhD,‡ Naoto Tsuchiya, PhD,§ Takashi Kohno, PhD,§ Vidar Skaug, MD,‖ Steen Mollerup, MD,‖ Aage Haugen, MD,‖ Paul S. Meltzer, MD, PhD,‡ Jun Yokota, MD,¶ Yae Kanai, MD, PhD,† and Curtis C. Harris, MD**

J. Thorac. Oncol. 10, 1037–1048 (2015).

A report on a validated, multiple ‘omics’ biomarker panel for lung adenocarcinoma (ADC).

· Background

Up to 30% stage I lung cancer patients suffer recurrence within 5 years of curative surgery. They sought to improve existing protein-coding gene and microRNA expression prognostic classifiers by incorporating epigenetic biomarkers.

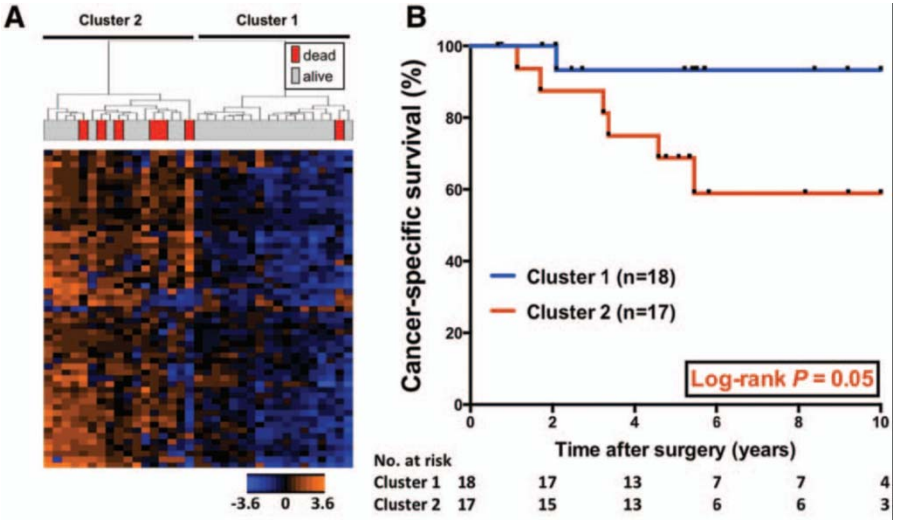
· Methods

Genome-wide screening of DNA methylation and pyrosequencing analysis of HOXA9 promoter methylation were performed in two independently collected cohorts of stage I lung adenocarcinoma.

Then, the prognostic value of HOXA9 promoter methylation alone and in combination with mRNA and miRNA biomarkers was assessed by Cox regression and Kaplan–Meier survival analysis in both cohorts.

- **HOXA9 promoter methylation stratifies lung cancer outcome in two independent patient cohorts**

National Cancer Institute (NCI) microarray cohort



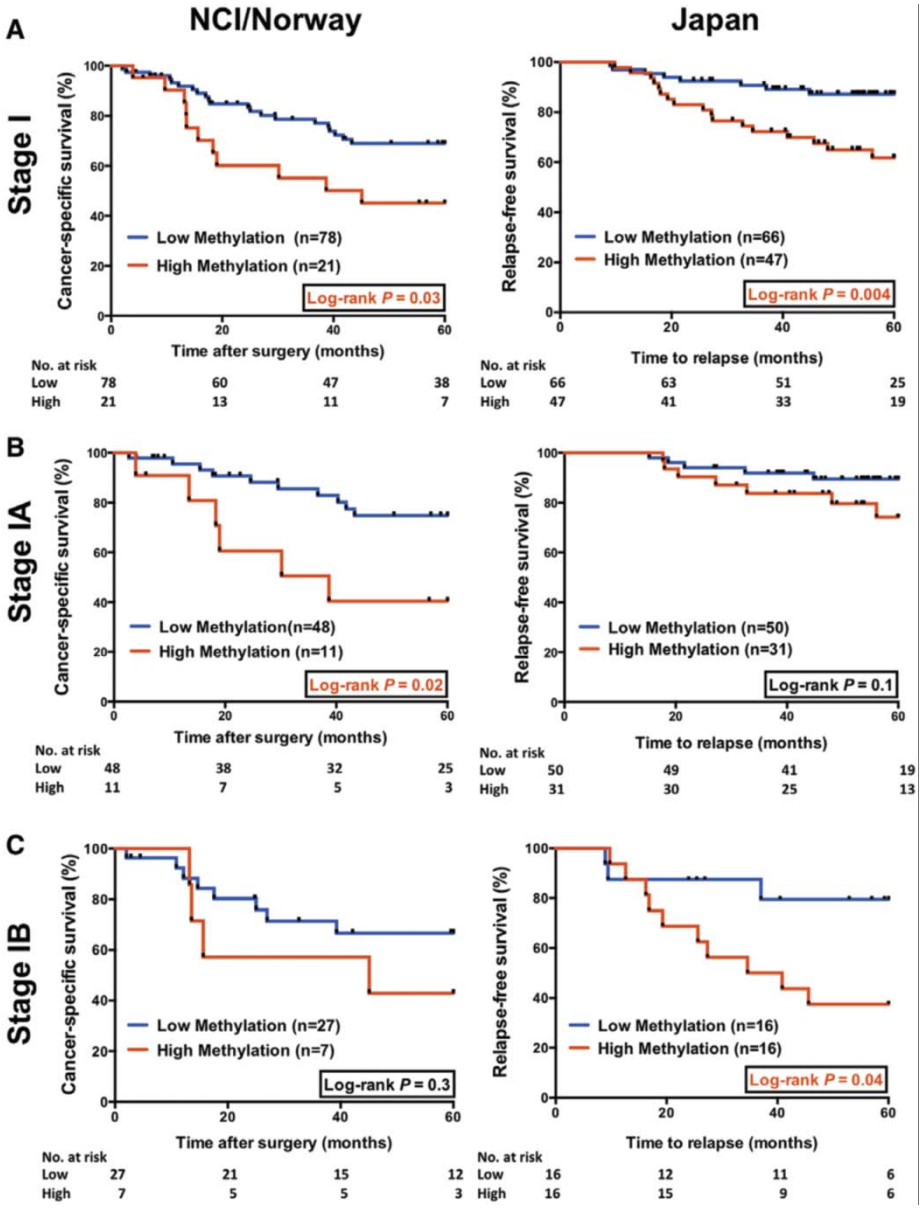
Japan microarray cohort

Each column represents an individual patient and each row an individual CpG probe.

55 probe sets (corresponding to 47 genes) that were hypermethylated in tumors.

Patients in the high methylation clusters had shorter cancer-specific (B) or relapse-free (D) survival in Kaplan–Meier survival analysis.

- **HOXA9 promoter methylation stratifies lung cancer outcome in two independent patient cohorts**

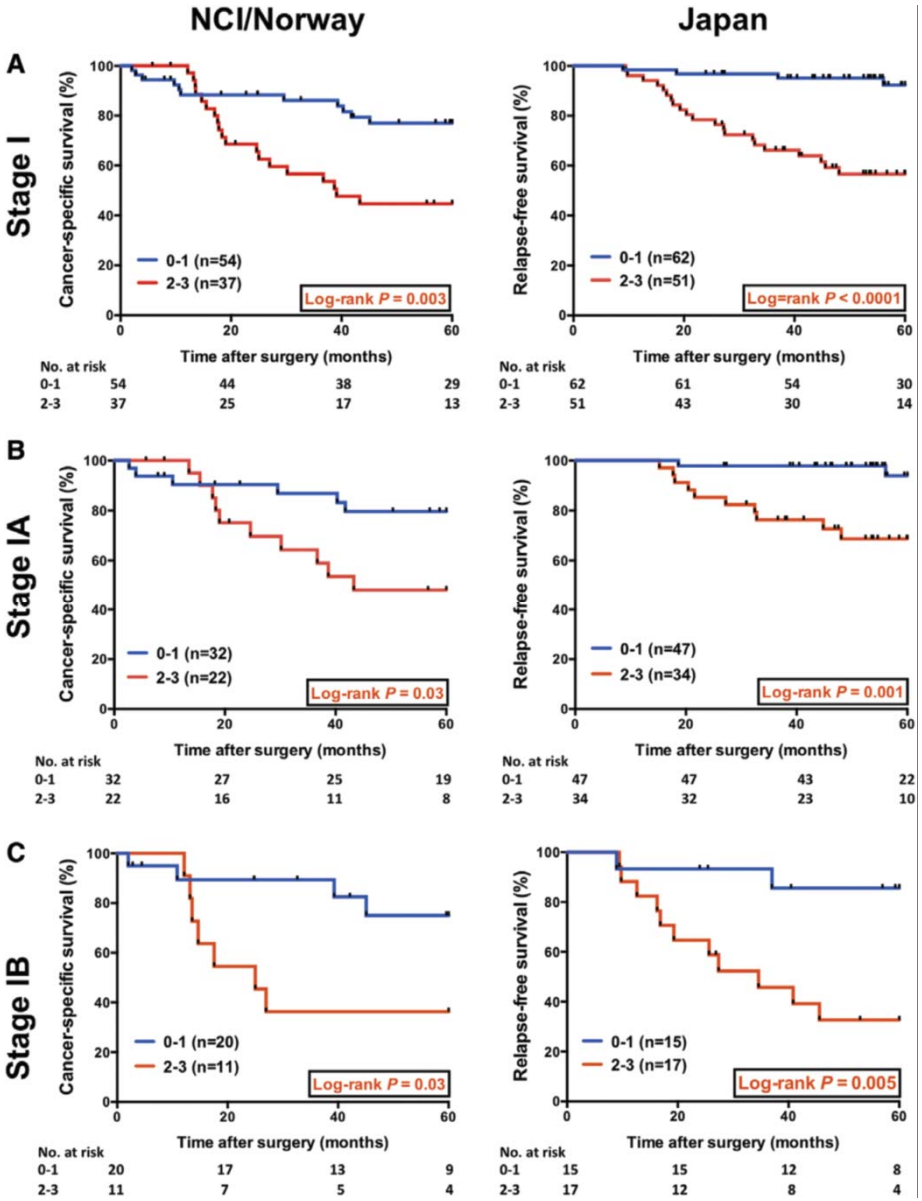


Kaplan–Meier survival analysis of HOXA9 promoter methylation in stage I lung adenocarcinoma.

- 3 probes were associated with HOXA9.
- HOXA9 methylation values were dichotomized based on ≥40% or <40% mean methylation in pyrosequencing analysis.

High HOXA9 methylation (above 40%) was observed in 23 of 104 (22%) stage I lung adenocarcinoma and associated with shorter cancer specific survival in the Kaplan–Meier and Cox regression analyses, independent of stage (IB versus IA) and smoking (table not shown).

- **A combination of RNA and DNA biomarkers identifies patients with poor prognosis in two independent cohorts of stage I lung ADC**



They previously developed **RNA-based prognostic biomarkers** that stratify stage I lung ADC patients. A combined high miR-21 and 4-protein-coding gene classifier (based on expression of XPO1, BRCA1, HIF1α, and DLC1) identified a subset of stage I lung ADC patients with poor prognosis in the same NCI/Norway and Japan cohorts.

They categorized patients according to the number of combined high values of HOXA9 promoter methylation, miR-21, and 4-protein-coding gene signature.

An increasing combined score conferred greater risk for poor outcome in stage I, and within subgroup analysis of stage IA and stage IB.

Conclusion of the paper

- A prognostic classifier comprising three types of genomic and epigenomic data (mRNA, microRNA, and DNA Methylation) may help guide the postoperative management of stage I lung cancer patients at high risk of recurrence.

- Their exploration of the lung cancer methylome in relation to gene and miRNA expression contributes to the molecular taxonomy of lung cancer and may have therapeutic implications.

- Their approach exemplifies the power of precision medicine to harness diverse molecular data to better categorize disease and inform treatment.

Thank you !

Wish you all the best for your future life !

Cover of Nature review cancer
August 2016 Vol 16 No 8