

Epigenetic markers of smoking-induced Lung Cancer: screening and risk

Florence Guida, Therese Nøst

Paolo Vineis, Torkjel Sandanger, Marc Chadeau-Hyam

Background

- ◇ Lung cancer is the first cause of death from cancer worldwide
- ◇ Survival rates remain low
 - 5 years : 10-20%
 - increase with the stage of lung cancer
- ◇ **Lack of reliable markers enabling an early diagnosis**

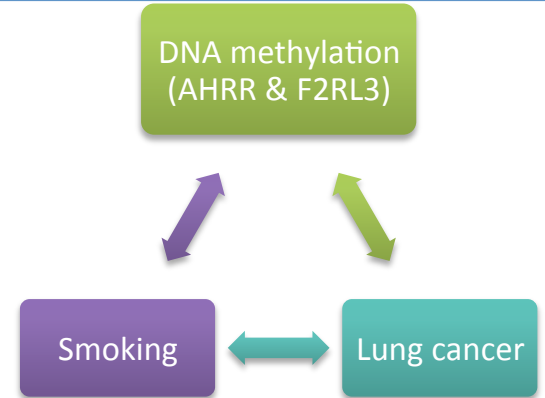
Background

- ◆ DNA methylation could mediate the effect of smoking on lung cancer
- ◆ Carcinogenic effect of tobacco smoking persists decades after smoking cessation

Vineis P et al., J Natl Cancer Inst, 2004

- ◆ Methylation dynamics after smoking cessation

- Identification of reversible/persistent methylation markers of smoking
- **Do they yield different risk profiles?**



Fasanelli et al., Nature communication, 2015



Human Molecular Genetics, 2015, Vol. 24, No. 8 2349-2359

doi: 10.1093/hmg/ddu751
Advance Access Publication Date: 2 January 2015
Original Article

ORIGINAL ARTICLE

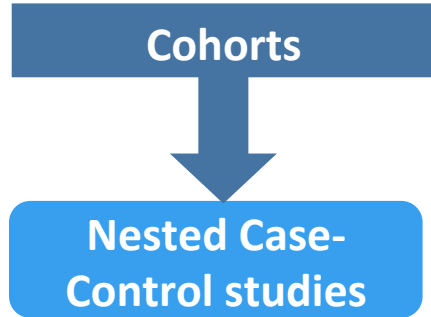
Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation

Florence Guida^{1,†}, Torkjel M. Sandanger^{2,†}, Raphaële Castagné^{1,†}, Gianluca Campanella¹, Silvia Polidoro³, Domenico Palli⁴, Vittorio Krogh⁵, Rosario Tumino⁶, Carlotta Sacerdote³, Salvatore Panico⁷, Gianluca Severi^{3,8,9}, Soterios A. Kyrtopoulos¹⁰, Panagiotis Georgiadis¹⁰, Roel C.H. Vermeulen^{1,11,12}, Eiliv Lund², Paolo Vineis^{1,3,†} and Marc Chadeau-Hyam^{1,11,†,*}

Guida et al., Hum. Mol. Gen., 2015

I – DNA methylation biomarkers of smoking

Population



EPIC-Italy
N=47,749

*European Prospective Investigation
into Cancer and Nutrition*

EPIC-Italy Lung
controls
N=85

EPIC-Italy Breast
& Colon
N=451

NOWAC
N=50,000

Norwegian Women and Cancer Study

NOWAC-Breast
N=336

NOWAC-Lung
controls
N=127

**999
Women**

- Blood samples at recruitment
- All cases have a time to diagnosis >1 year

Methods

◇ Aim

- to identify for each probe the the time since smoking cessation (t) where significance is lost

◇ Binary smoking status recoded as a function of time since quitting smoking (t)

- Principle: pool (i) long-term former smokers and never smoker, and (ii) short-term former and current smokers
- For $t = 0$: current smokers vs non smokers at inclusion, and $t = t_{\max}$ never vs ever smokers
- For each value of t we run an EpWAS
- the value of t after which pooling yielded too heterogeneous populations to detect smoking-induced signals.

Non smokers
(Reference)

Never smokers &
Former smokers

Never smokers &
Former smokers $t \geq x$ years

Never smokers

0

x

t_{\max}

t

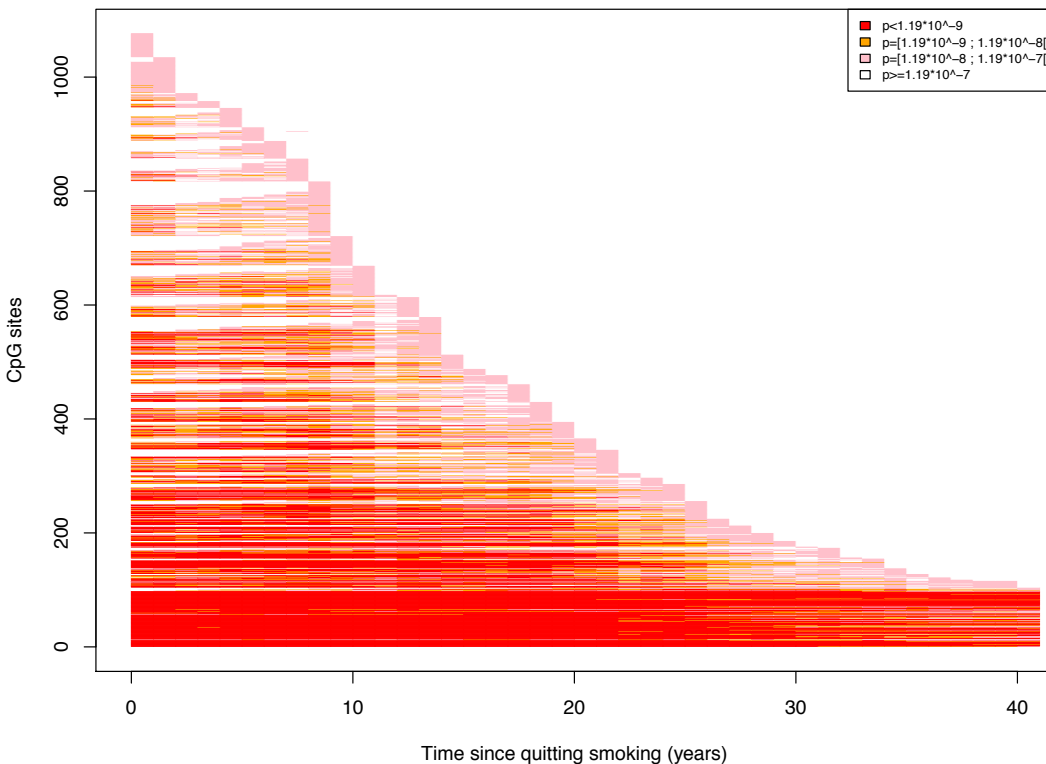
Smokers

Current smokers

Current smokers &
Former smokers $t < x$ years

Current smokers &
Former smokers

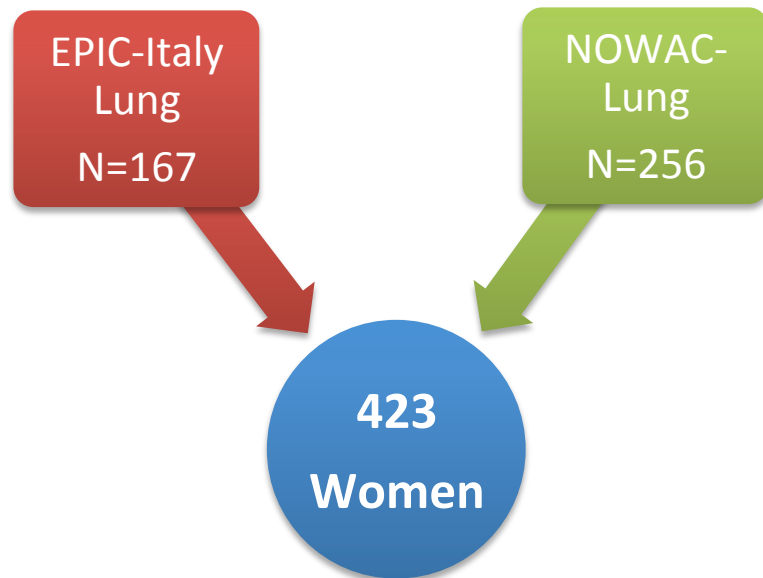
I-1. Significance trajectories



- ◇ Generalised linear model
 - **Methylation beta-values** ~ **smoking status + case/control status**
 - Adjustment for technical covariates and matching variables
 - ◇ 1076 associations at least once
 - ◇ To objectively classify those CpGs sites : k-means clustering
 - **Reversible** biomarkers N=730
- Methylation reverts back to normal after smoking cessation
- **Persistent** biomarkers N=346
- Still differentially methylated >30 years after smoking cessation

II – Link of those smoking related methylation markers with lung cancer

Population



	Cases		Controls	
	N	%	N	%
	mean	SD	mean	SD
Total	211		212	
Age at sampling	56	5.9	56	5.9
Time to diagnosis	5.24	3.23	-	-
Histological types				
Adenocarcinomas	101	47.87	-	-
Large-cell carcinoma	18	8.53	-	-
Small-cell carcinoma	34	16.11	-	-
Squamous-cell carcinoma	28	13.27	-	-
Other lung cancers	30	14.22	-	-
Smoking status				
Never	35	16.59	104	49.06
Former	51	24.17	51	24.06
Current	125	59.24	57	26.89
Cohort				
EPIC-Italy	82	38.86	85	40.09
NOWAC	129	61.14	127	59.91

Targeted analysis

◇ For the 1076 CpGs associated with smoking

- Unconditional logistic regression
 - Case/control ~ methylation
 - Adjustment for technical covariates + matching criteria

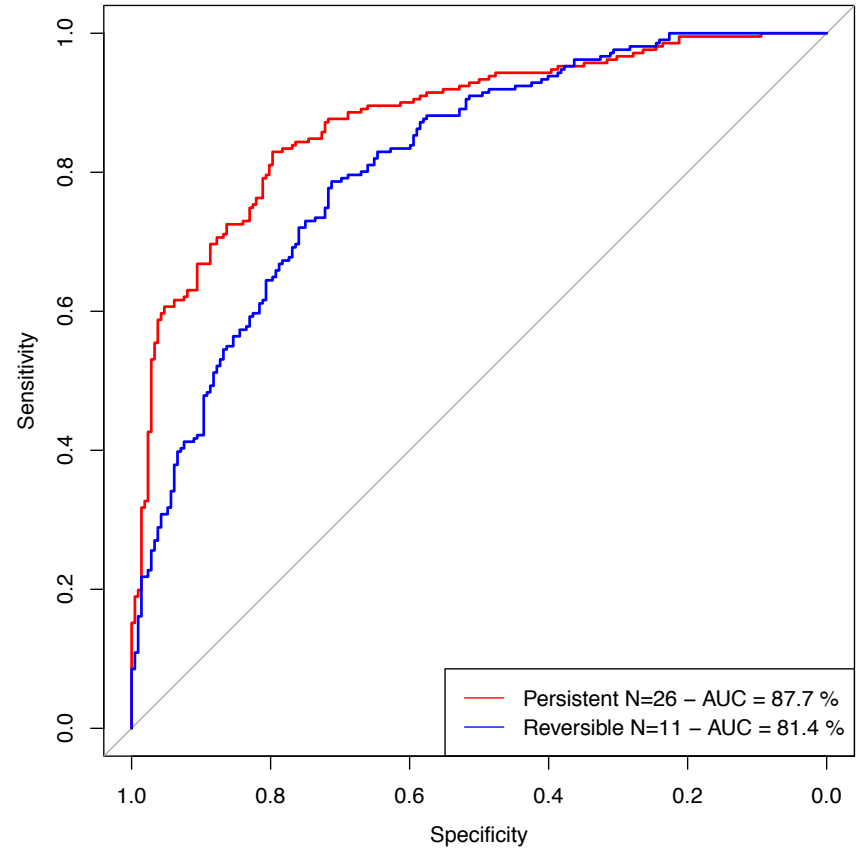
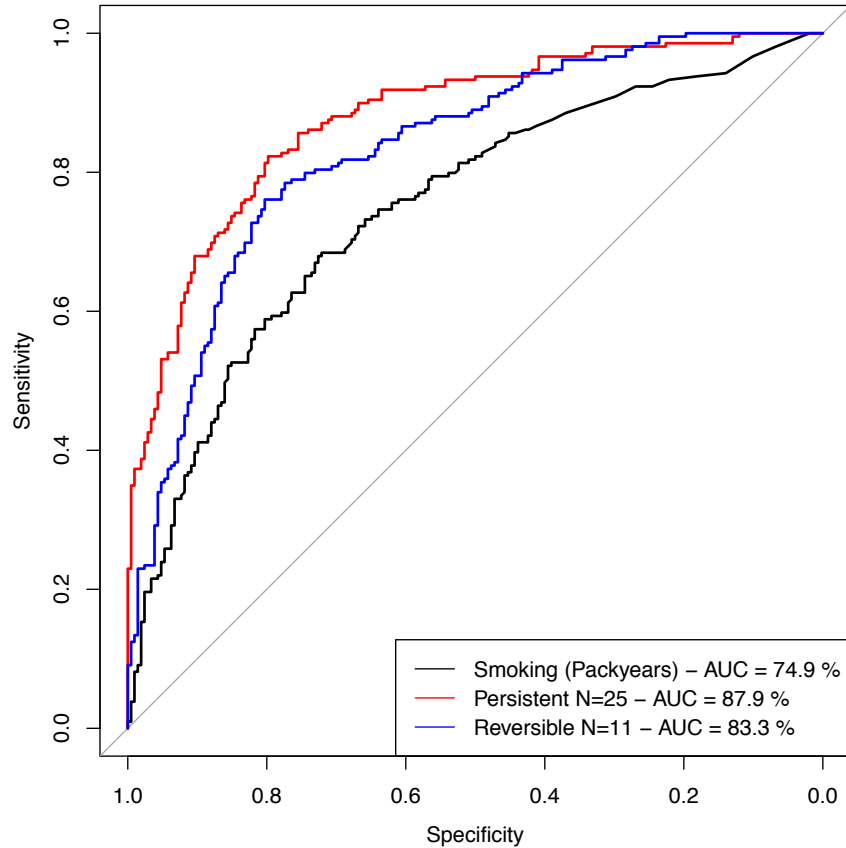
◇ 157 out of 1076 significant CpGs (Bonferroni level)

	Total	$\beta > 0$	$\beta < 0$
Persistent	111 (32%)	1	110
Reversible	46 (6%)	1	45

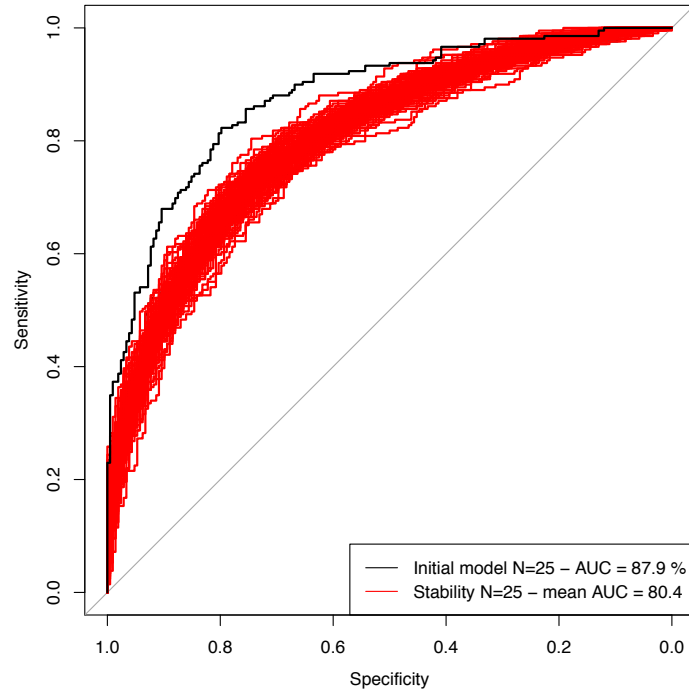
◇ Stepwise procedure (backward)

- To prioritise the most relevant CpG sites
 - 25(/111) persistent CpGs
 - 11(/46) reversible CpGs

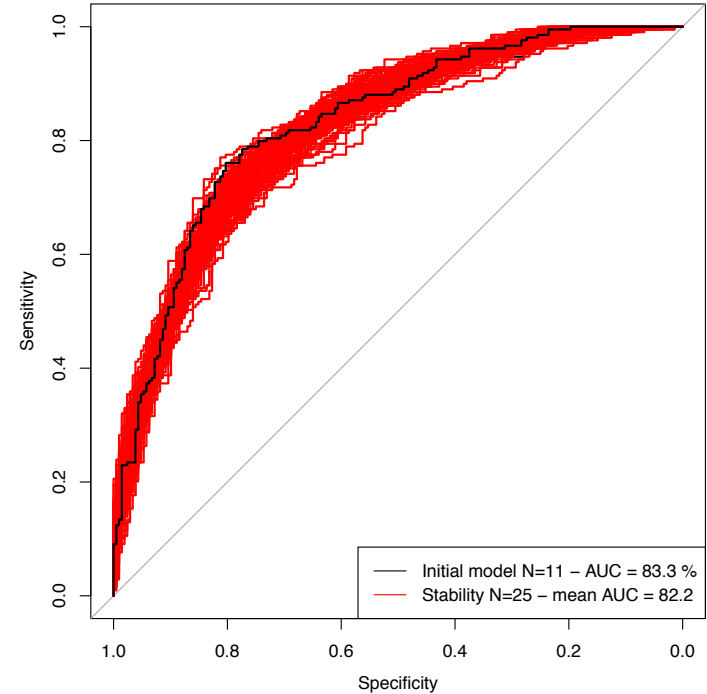
ROC Curves



Sensitivity analyses : persistent sites

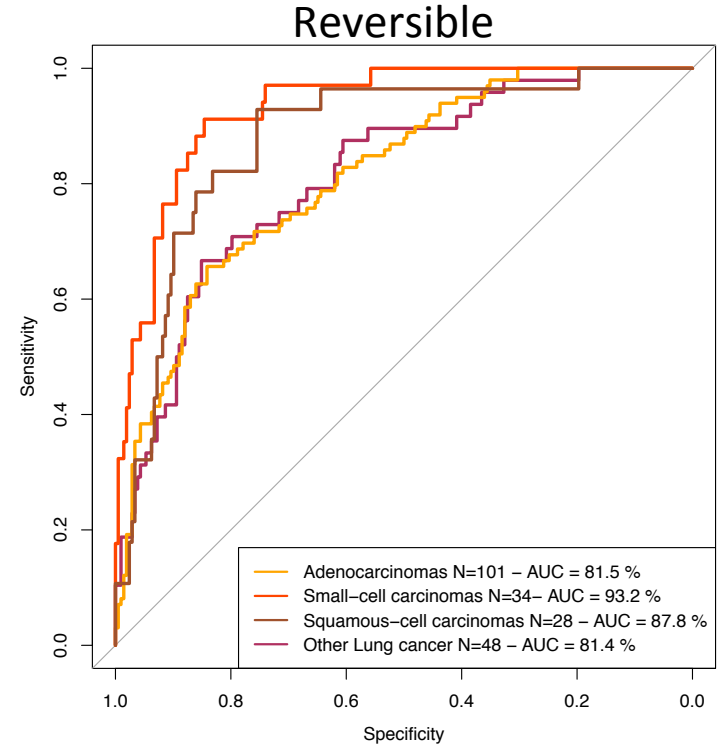
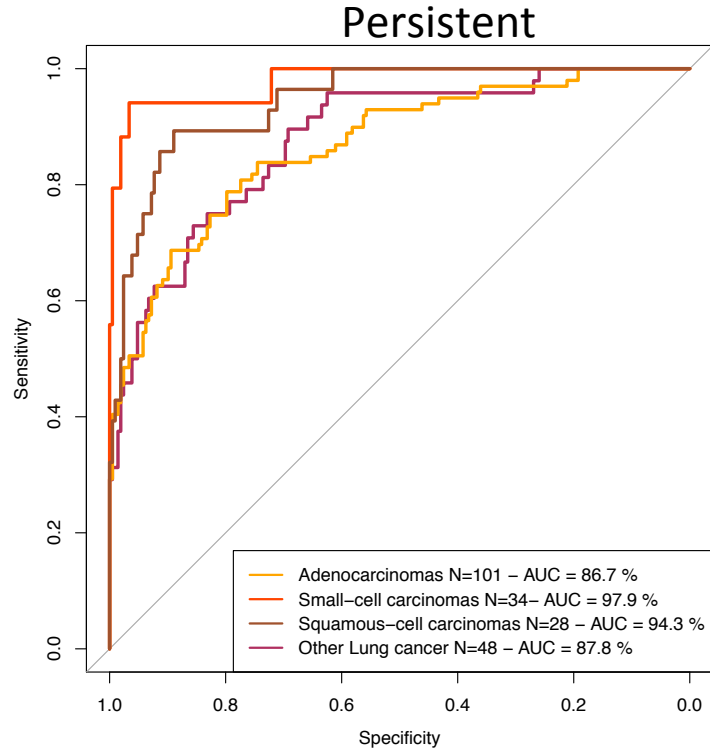


◆ Influence of the number of selected CpG sites



◆ Relevance of the 25 selected CpG sites

Analyses by histological type of lung cancer



- Warning: non-cross validated AUC estimates (ongoing; to rule out overfitting)
- Subtype stratified analyses comprise few cases

Conclusion & Perspectives

- ◆ Methylation biomarkers have the potential to improve lung cancer risk prediction
 - Persistent markers seem to have a better predictive ability than reversible markers
 - ➔ additional sensitivity analyses are ongoing: persistent CpG sites are those mostly affected by smoking
- ◆ Apply the same methodology to gene expression data
 - Assess predictive ability of GE only
 - GE x Methylation integration:
 - Functional interpretation of differentially CpGs
 - Assess joined contribution to prediction

Acknowledgments

- ◆ Paolo Vineis
- ◆ Marc Chadeau-Hyam
- ◆ Gianluca Campanella
- ◆ Michelle Plusquin

**Imperial College
London**

- ◆ Therese H Nøst
- ◆ Torkjel M Sandanger
- ◆ Eiliv Lund



◆ **Funders**

- ◆ *The COLT Foundation*
- ◆ *MRC-PHE Centre for Environment & Health*
- ◆ *EU project –Exposomics*

- ◆ Paolo Vineis
- ◆ Silvia Polidoro



- ◆ Roel Vermeulen



Universiteit Utrecht

- ◆ Mattias Johansson

[International Agency for Research on Cancer](#)

