Cell Line Genomic Profiling and Neuroendocrine Expression Signature in Lung Cancer (1U01CA213338)

Luc Girard, Ph.D.
Assistant Professor of Pharmacology
Hamon Center for Therapeutic Oncology Research
UT Southwestern Medical Center

SCLC Research Consortium
March 15th, 2018

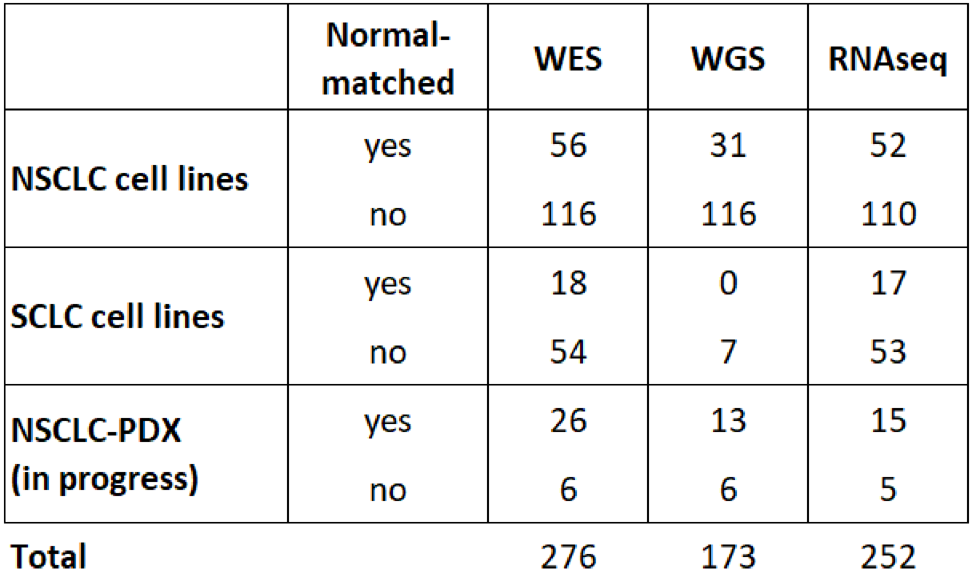
1) Lung Cancer Cell Line Genomic Profiling

2) A Neuroendocrine Expression Signature for Lung Cancer

Sequencing of Lung Cancer Cell Lines

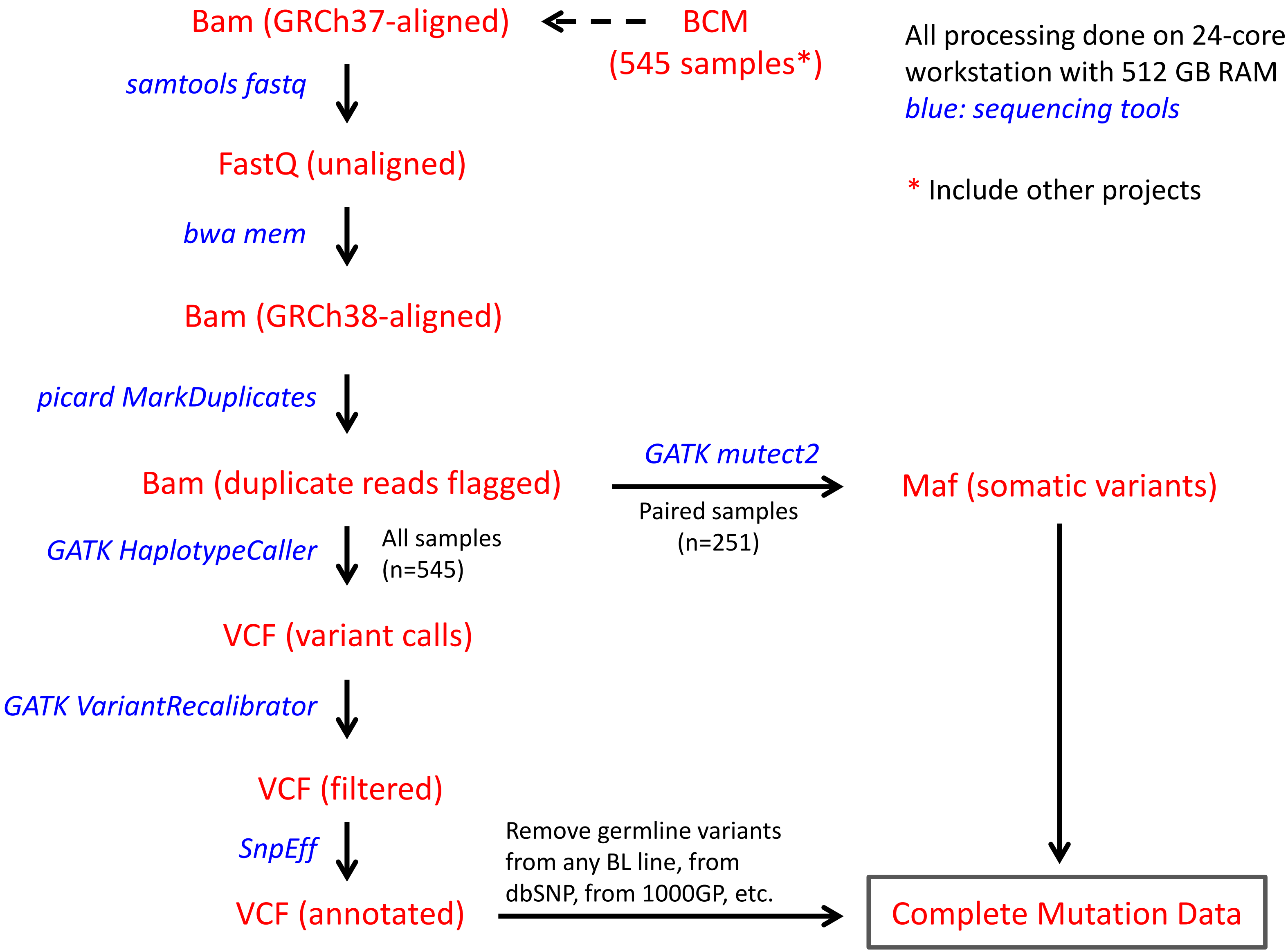
- DNA and RNA from lung cancer cell lines and PDXs (and when available germline DNAs) were prepared at UTSW (Kenneth Huffman, Boning Gao, Brenda Timmons) and sent to David Wheeler at Baylor College of Medicine (BCM) for sequencing as part of CPRIT MIRA (RP110708-P1) grant

- Whole Exome Sequencing (WES), Whole Genome Sequencing (WGS) and RNAseq were obtained as follows:



	Normal-matched	WES	WGS	RNAseq
NSCLC cell lines	yes	56	31	52
	no	116	116	110
SCLC cell lines	yes	18	0	17
	no	54	7	53
NSCLC-PDX (in progress)	yes	26	13	15
	no	6	6	5
Total		276	173	252

Exome Sequencing of Lung Cancer Cell Lines and PDXs



Bam (GRCh37-aligned) ← BCM (545 samples*)

samtools fastq

FastQ (unaligned)

bwa mem

Bam (GRCh38-aligned)

picard MarkDuplicates

Bam (duplicate reads flagged)

GATK HaplotypeCaller

All samples (n=545)

VCF (variant calls)

GATK VariantRecalibrator

VCF (filtered)

SnpEff

VCF (annotated)

Remove germline variants from any BL line, from dbSNP, from 1000GP, etc.

Bam (duplicate reads flagged) → *GATK mutect2* → Maf (somatic variants)

Paired samples (n=251)

Complete Mutation Data

All processing done on 24-core workstation with 512 GB RAM

blue: sequencing tools

* Include other projects

Most Frequently Mutated Cancer* Genes in SCLC Cell Lines and Tumors

Mutation Frequencies

18 Paired SCLC Lines
(avg mut count = 17)

54 Unpaired SCLC Lines
(avg mut count = 34)

Gene	George (SCLC Tumors)	Paired SCLC Lines	Unpaired SCLC Lines	H1672	H2141	H2227	H1450	H289	HCC2433	HCC4003	H1184	H128	H2195	H2107	HCC4001	H1339	H1963	HCC33	H209	H2171	HCC970
TP53	94%	77%	83%	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	
RB1	79%	64%	56%	M	M	M	M		M		M	M	M	M	M		M	M	M	M	
KMT2D	0%	27%	38%	M			M						M			M			M	M	
FAM135B	24%	23%	19%		M	M				M							M	M			
FAT3	20%	18%	39%	M	M		M										M				
FAT1	15%	18%	22%	M			M	M							M						
RHOA	0%	18%	2%	M								M	M						M		
FAM47C	9%	18%	13%			M						M		M	M						
CREBBP	10%	18%	14%					M		M		M						M			
FAT4	11%	18%	20%	M	M							M			M						
CDH10	16%	5%	16%										M								
NOTCH1	15%	5%	16%			M															
CDH11	4%	14%	6%	M		M													M		
CTNNA2	5%	14%	20%						M	M		M									
PTPRT	10%	14%	16%							M			M						M		
KIAA1549	4%	14%	11%									M	M			M					
NTRK3	4%	14%	8%		M	M														M	
AKAP9	5%	14%	14%											M	M	M					
ROS1	8%	14%	19%	M		M		M													
ZEB1	9%	14%	6%	M	M				M												
PTPRB	14%	14%	22%								M				M	M					
ZNF521	12%	9%	11%	M																M	
EP300	12%	9%	17%											M		M					
PREX2	11%	9%	11%	M								M									
RGS7	10%	9%	8%						M		M										
GRM3	10%	0%	3%																		
PDE4DIP	10%	0%	3%																		
AMER1	0%	9%	6%		M			M													
CBLB	2%	9%	6%				M				M										
EPHA3	3%	9%	5%				M									M					
ARID1A	4%	9%	14%					M				M									
PDGFRA	5%	9%	11%			M						M									
CTNND2	9%	9%	13%						M					M							
KMT2C	0%	9%	25%				M										M				
UBR5	3%	9%	9%								M				M						
ATR	5%	9%	9%						M		M										

Gene	H510	H1522	H1048	H1514	H740	H250	H748	H446	H1930	H1436	H345	H524	H60	H774	HCC4004	SHP-77	H1836	H1092	H146	H1607	H211	DMS-153	HCC4000	H1688	H187	H711	H1876
TP53	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
RB1	M		M		M	M	M	M	M			M	M	M	M				M	M		M	M			M	M
KMT2D	M	M	M	M			M	M	M	M			M		M	M			M		M	M	D				M
FAM135B					M				M				M	M			M	M	M		M						
FAT3	M		M		M		M	M		M	M		M		M	M	M		M								M
FAT1	M	M	M					M									M			M			M				M
RHOA							M																				
FAM47C							M								M					M			M			M	
CREBBP		M	D	M							M																M
FAT4			M	M	M		M		M	M					M		M							M		M	
CDH10								M			M													M			
NOTCH1										M	M											M				D	
CDH11																M	M						M				
CTNNA2	M		M	M	M			M	M				M	M										M			
PTPRT					M									M								M					M
KIAA1549	M		M			M																					
NTRK3		M												M			M	M				M					
AKAP9			M		M	M						M	M	M			M										
ROS1					M		M					M	M								M		M			M	
ZEB1						M					M										M						
PTPRB	M				M			M		M				M	M			M			M						M
ZNF521							M		M		M													M			
EP300		M			M		M					M					M			M			M		M		M
PREX2															M		M					M		M			
RGS7	M					M								M													
GRM3																											M
PDE4DIP											M																
AMER1	M													M						M							
CBLB																					M			M			
EPHA3																M											
ARID1A	D		M				D			M				D		M										M	
PDGFRA				M				M													M						M
CTNND2						M			M			M	M	M													
KMT2C				M									M	M	M	M		M				M		M	M		
UBR5								M	M	M												M				M	
ATR			M			M						M					M										

Gene	H1926	H526	H69	H82	H1417	H1882	H2081	H3282	H2029	DMS-79	H1284	H1238	H738	H378	H735	H1694	H1870	H1994	HCC4005	H2679	H1628	H865	H841	H2198	H889	H2196	H2108
TP53	M	M	M	M	M	M	M	M	M	M	M	M	M		M	M	M	M	M	M	M	M	M	M	M	M	M
RB1	M	M	M	M	M	M	M		M	M	M	M			M	M	M	M	M					M		M	M
KMT2D						M	M	M	M		M		M											M		M	
FAM135B		M					M						M						M								
FAT3	M	M		M		M			M	M						M		M		M		M	M	M			
FAT1	M	M		M		M					M		M														
RHOA																											
FAM47C				M		M									M												
CREBBP						M					M					M			M								
FAT4									M									M				M					
CDH10		M							M			M						M	M					M		M	
NOTCH1		M		M						M	M		M								M						
CDH11							M																				
CTNNA2									M							M		M							M		
PTPRT						M	M						M									M		M		M	
KIAA1549													M		M									M		M	
NTRK3																											
AKAP9												M															M
ROS1		M	M							M				M								M					
ZEB1										M																	
PTPRB						M		M					M								M	M					
ZNF521					M					M															M		
EP300	M																										M
PREX2	M															M									M		
RGS7								M												M							
GRM3						M																					
PDE4DIP									M																		
AMER1																				M							
CBLB			M																						M		
EPHA3		M		M																							
ARID1A														D			D										
PDGFRA					M	M								M													
CTNND2	M														M												M
KMT2C								M				M	M							M	M	M	M				
UBR5												M															
ATR							M															M					

M Point Mutation

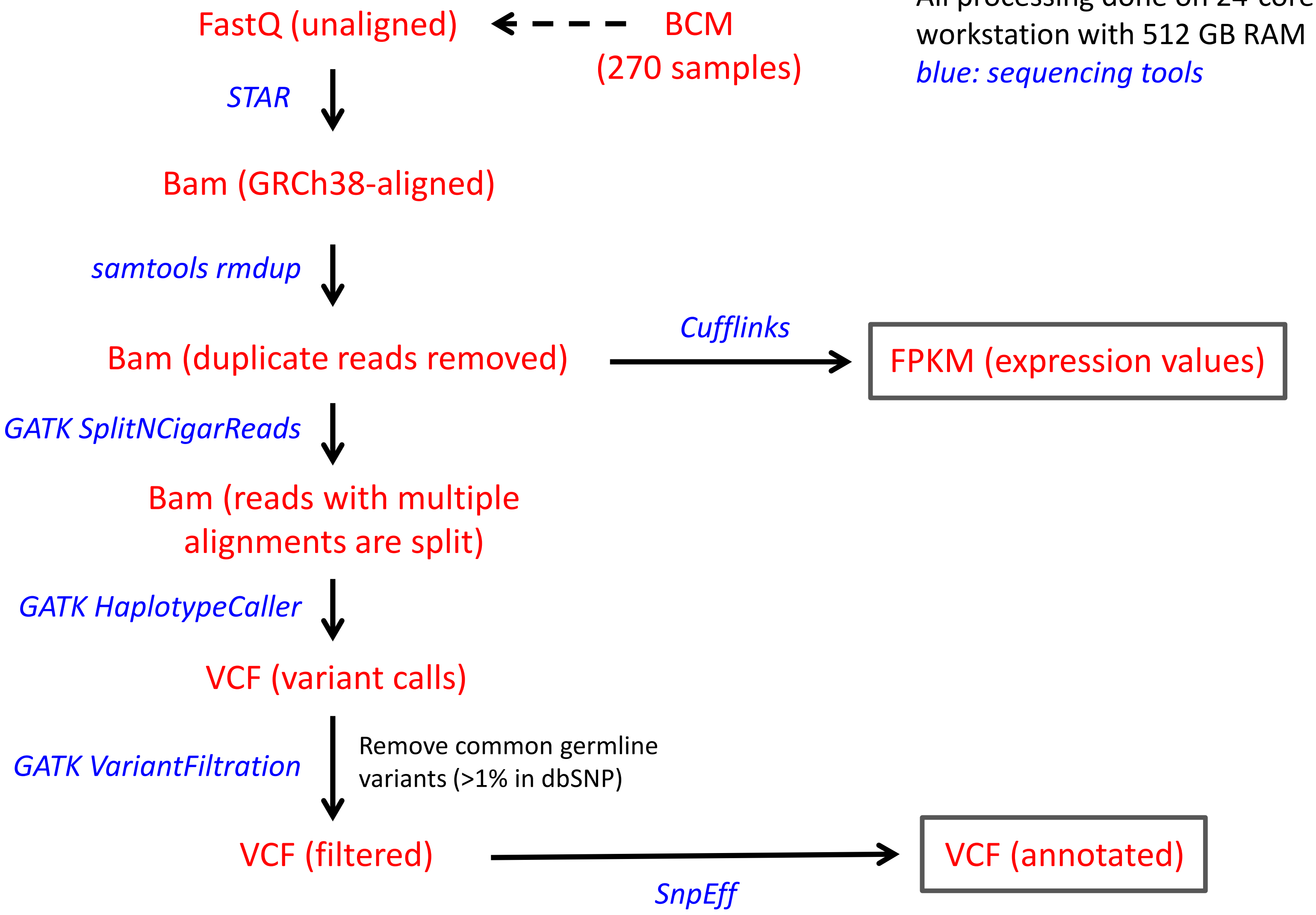
I Insertion

D Deletion

* Cancer Gene Census, 700 genes (COSMIC)

RNAseq of Lung Cancer Cell Lines and PDXs

All processing done on 24-core
workstation with 512 GB RAM
blue: sequencing tools



BCM
(270 samples)

FastQ (unaligned)

STAR

Bam (GRCh38-aligned)

samtools rmdup

Bam (duplicate reads removed)

Cufflinks

FPKM (expression values)

GATK SplitNCigarReads

Bam (reads with multiple
alignments are split)

GATK HaplotypeCaller

VCF (variant calls)

GATK VariantFiltration

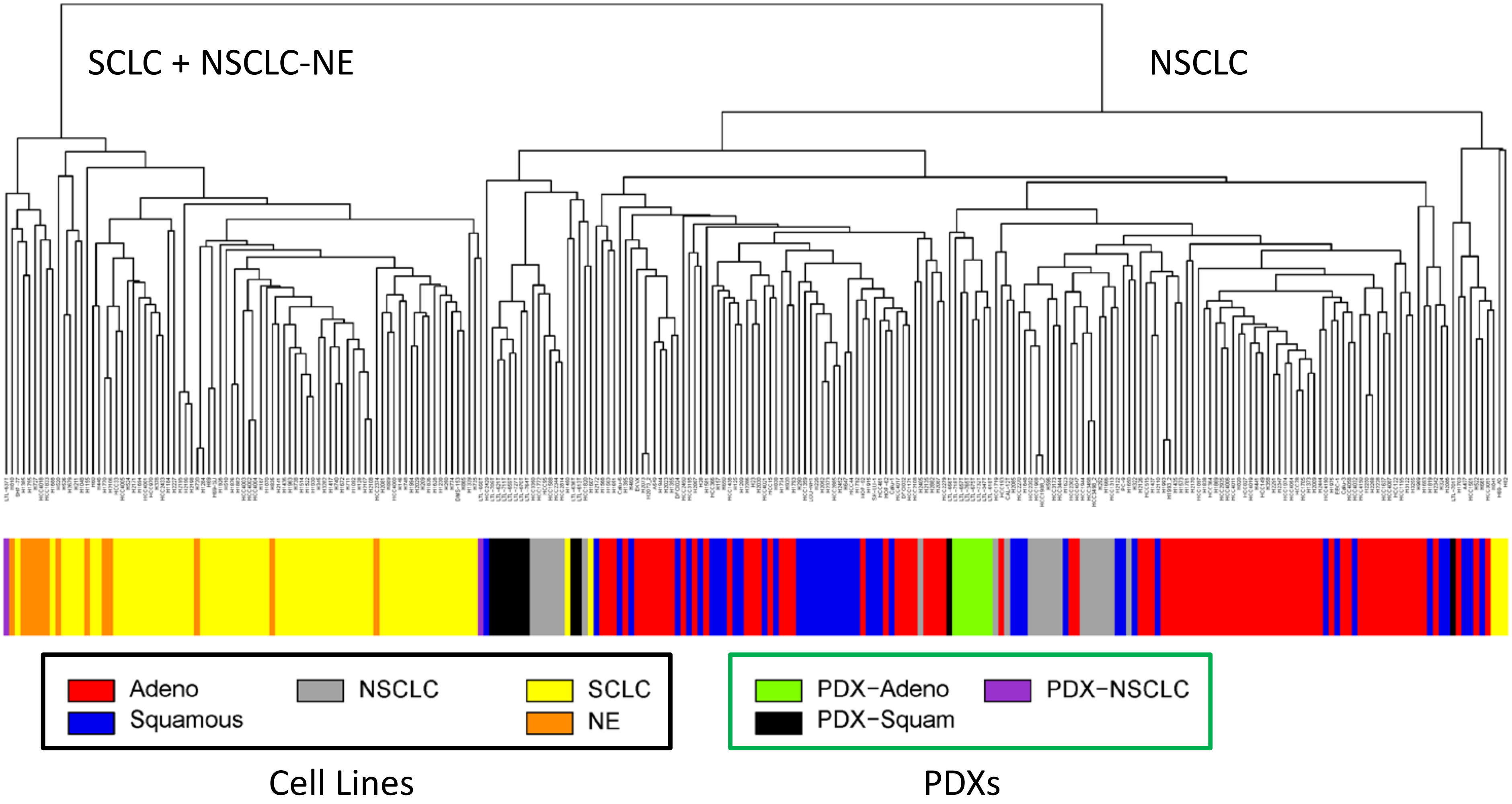
Remove common germline
variants (>1% in dbSNP)

VCF (filtered)

SnpEff

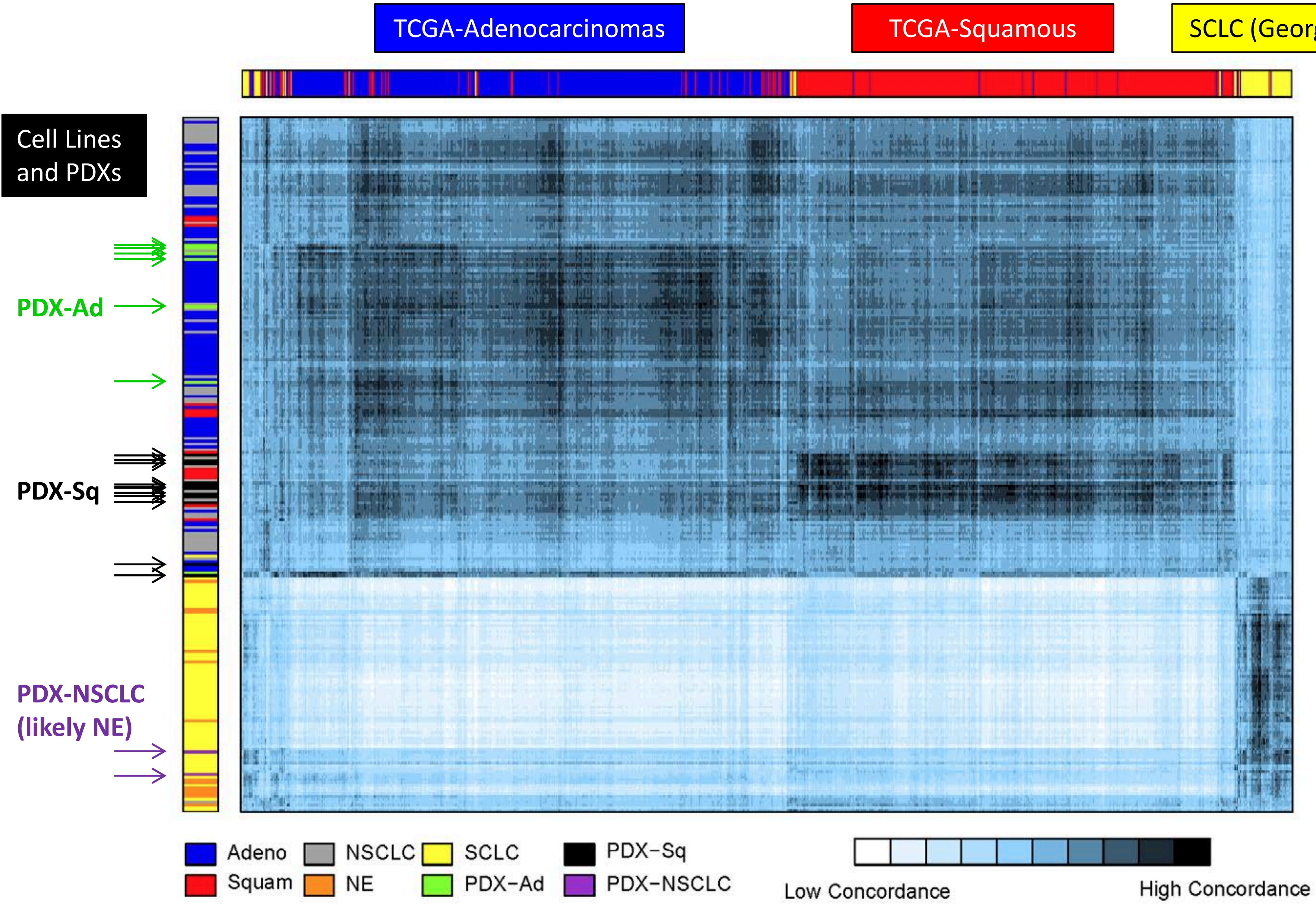
VCF (annotated)

Hierarchical Clustering – PDX and Cell Lines (RNAseq)



NE: Neuroendocrine
PDX: Patient-Derived Xenograft

2D RNAseq Clustering – PDXs & Cell Lines vs Primary Tumors



1) Lung Cancer Cell Line Genomic Profiling

2) A Neuroendocrine Expression Signature for Lung Cancer

SCLC is a Neuroendocrine Tumor

Definition of Neuroendocrine (NE) cells

- The production of a neurotransmitter, neuromodulator or neuropeptide hormone (similar to neuronal cells)
- The products are stored in membrane bound dense-core secretory granules from which the hormones are released by exocytosis in response to an external stimulus (similar to some neuronal cells)
- The absence of axons and synapses (unlike neurons)

Why do we need an NE signature?

- SCLCs display a range of neuroendocrine properties. A signature would help classify distinct subgroups and these may have different responses to therapy.
- Many NSCLCs display neuroendocrine properties. They are variously called carcinoids, large cell neuroendocrine, or even NSCLC-NOS.

Development of a Neuroendocrine Expression Signature

Adrenal cortex (endocrine) vs adrenal medulla (neuroendocrine)

↓ *Derive and apply adrenal NE signature*

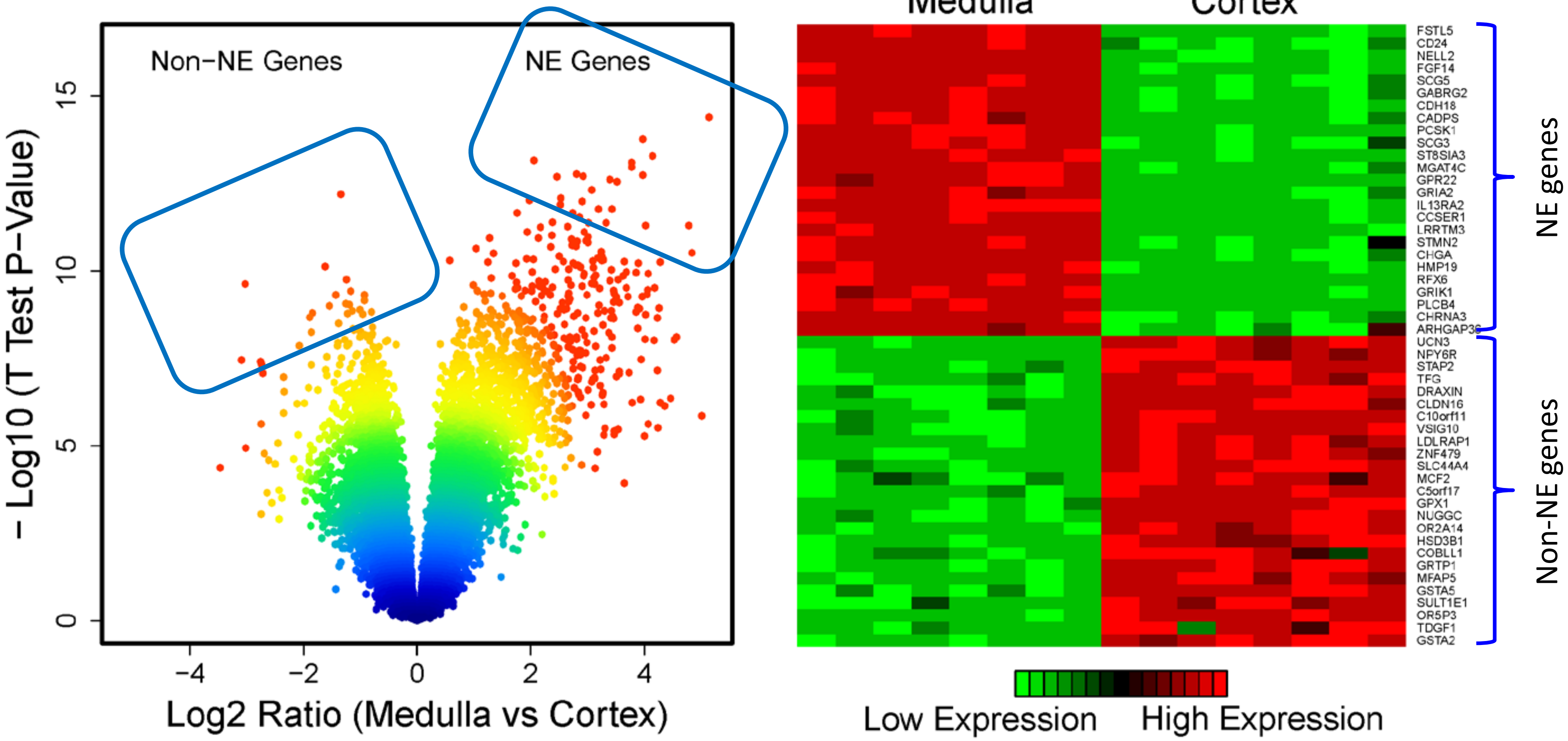
Identify NE and non-NE lung cancer cell lines

↓ *Derive and apply lung-specific NE signature*

Generate an NE score for lung cancer cell lines and tumor specimens

Development of a 50-Gene Adrenal NE Signature

Top 25 non-NE and top 25 NE genes (ranked by their distance to the origin) are selected

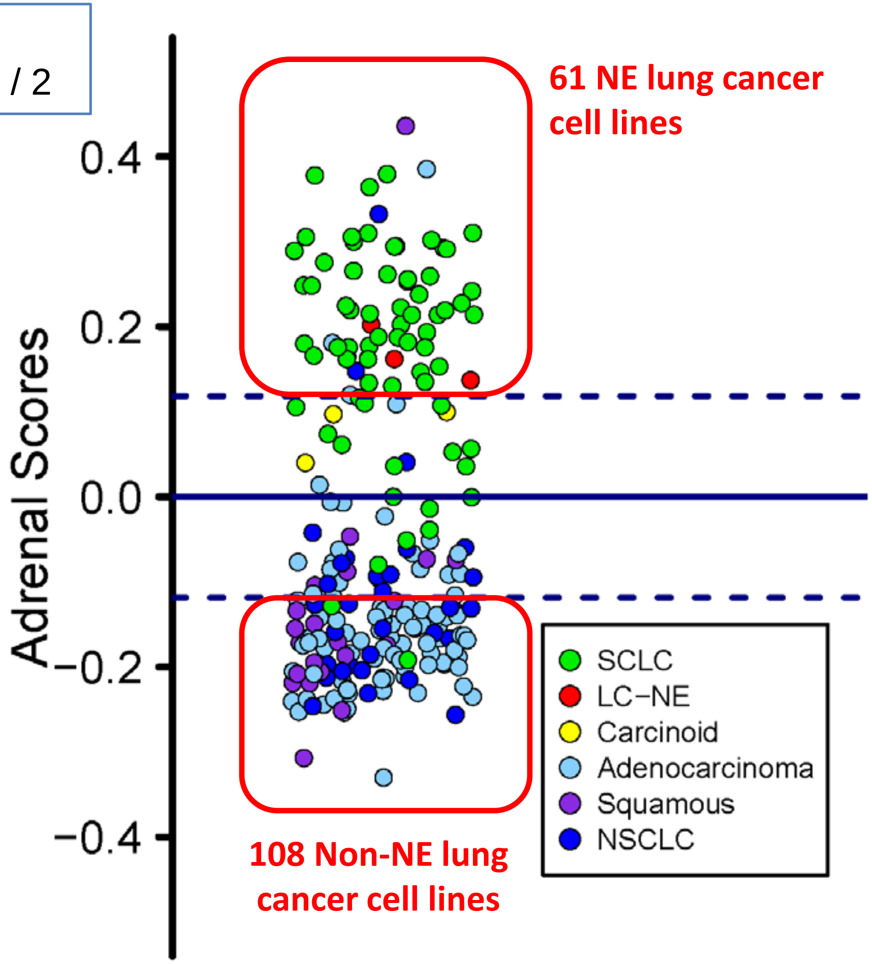


8 adrenal medulla vs 8 adrenal cortex non-malignant samples
(Affymetrix HuGene 1.0 microarrays; Karel Pacak, NIH/NICHD)

Adrenal NE Signature Applied to Lung Cancer Cell Lines

Adrenal NE Score =
(Correl NE – Correl Non-NE) / 2

Correl NE = Pearson correlation between expression of the 50 genes in the test sample and the average expression of the 50 genes in the medulla class of the training set



61 NE lung cancer cell lines

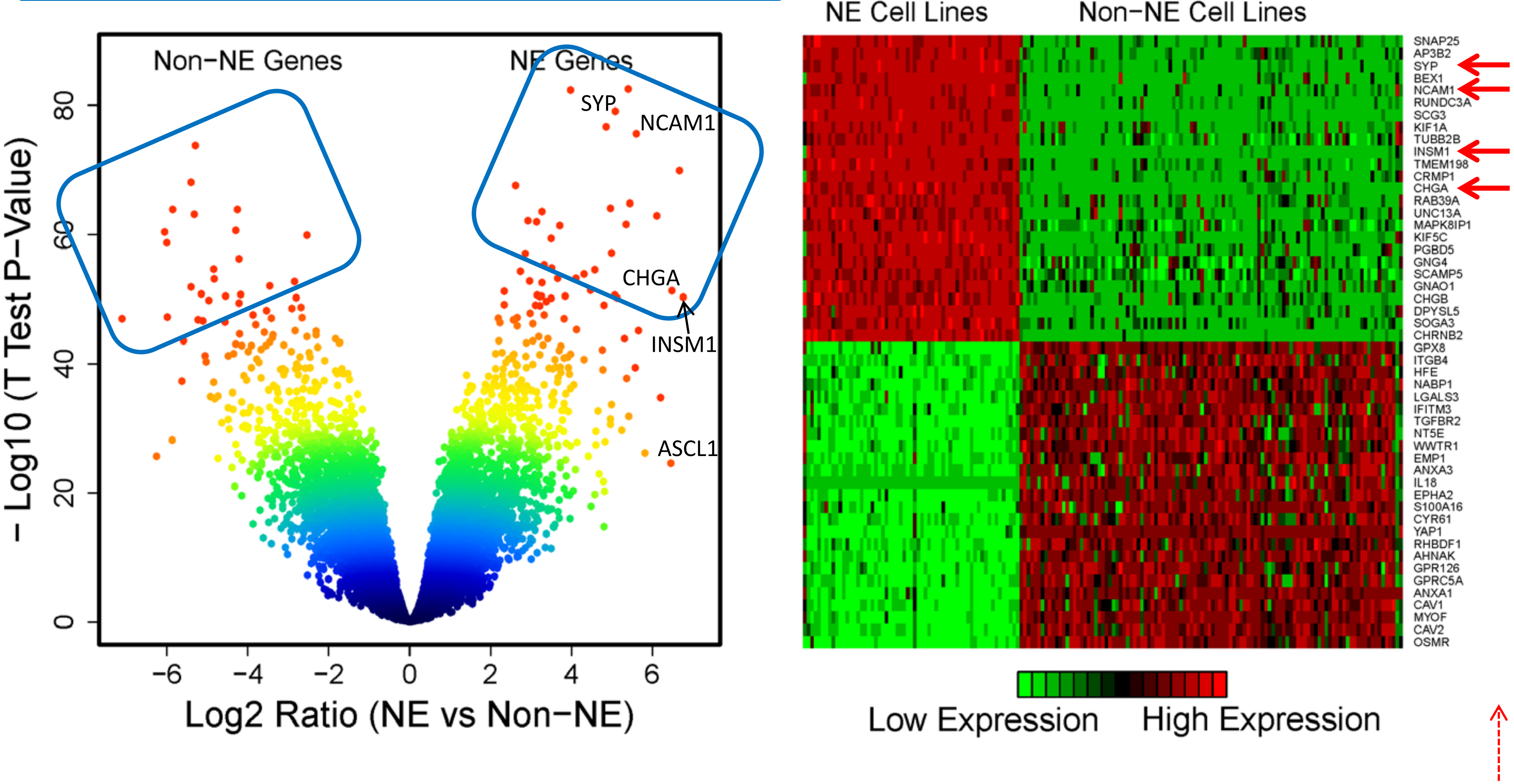
108 Non-NE lung cancer cell lines

Cutoff =
SD(abs(scores of training set))

RNAseq from 70 SCLC and 155 NSCLC cell lines

Development of a 50-Gene Lung-Specific NE Signature

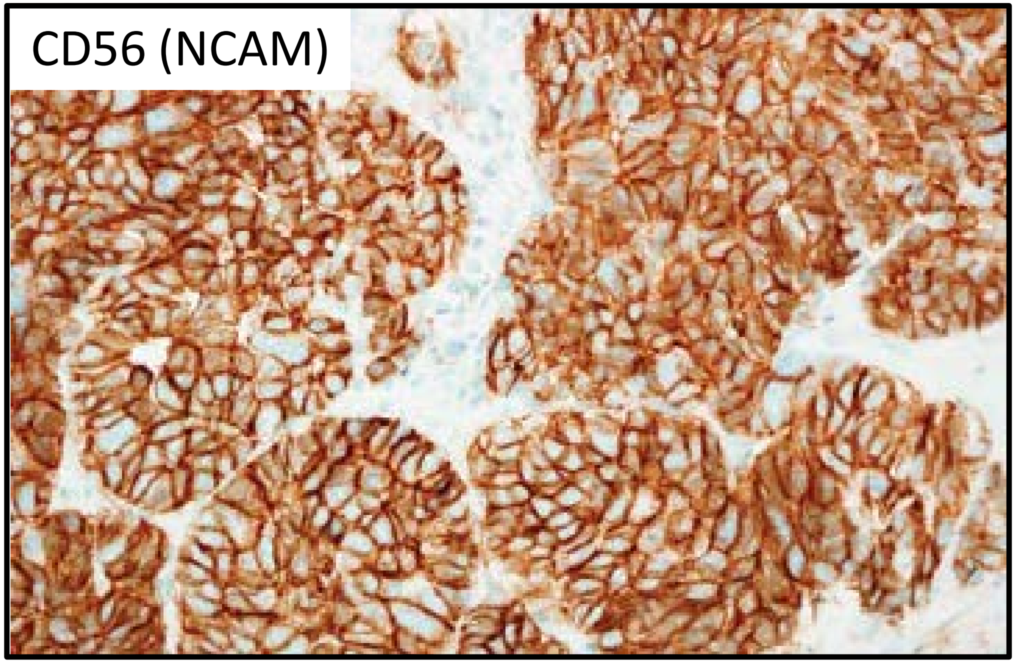
Top 25 non-NE and top 25 NE genes are selected

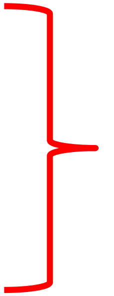


61 NE vs 108 non-NE lung cancer cell lines (RNAseq)

Arrows: Common immunostains

Immunostains in Common Pathological Use for Diagnosis of SCLC





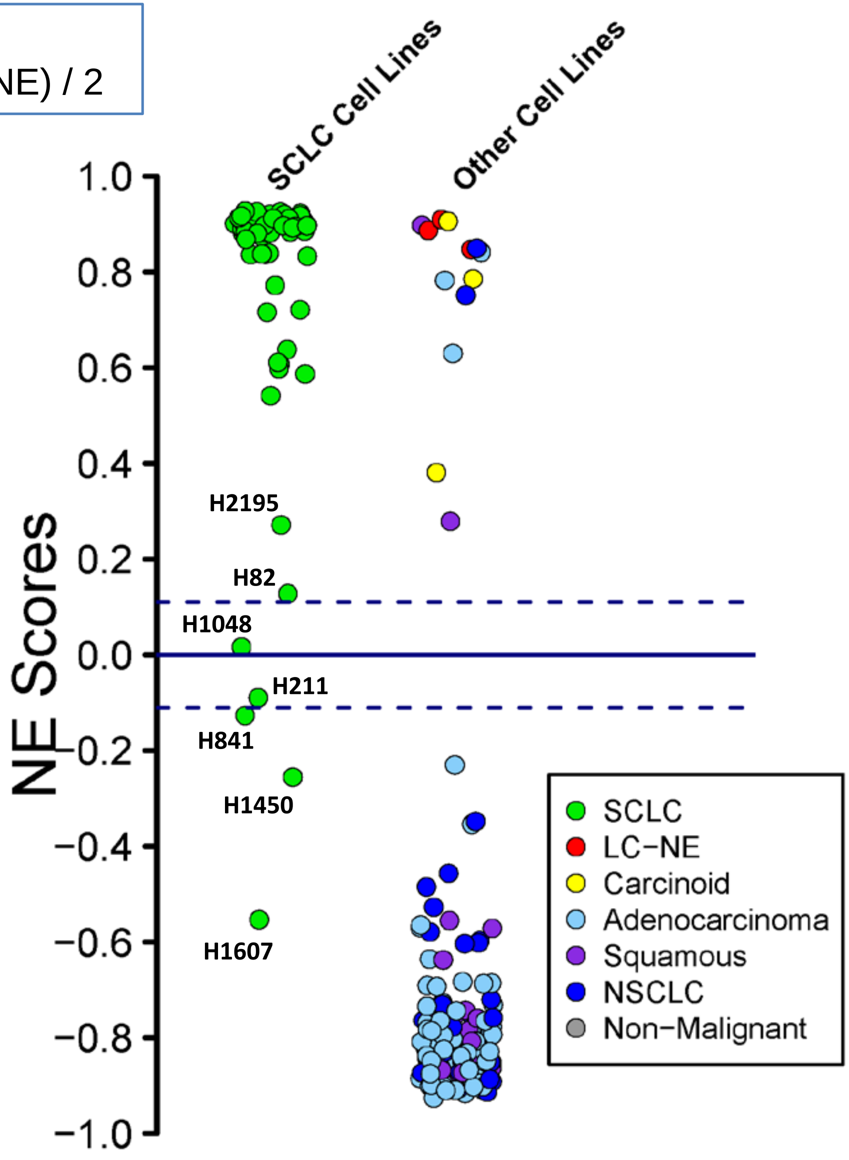
- Chromogranin A (CHGA)
- Synaptophysin (SYP)
- CD56 (NCAM)
- INSM1
- ASCL1 (emerging marker) →

Part of the NE signature (top 25 NE genes)

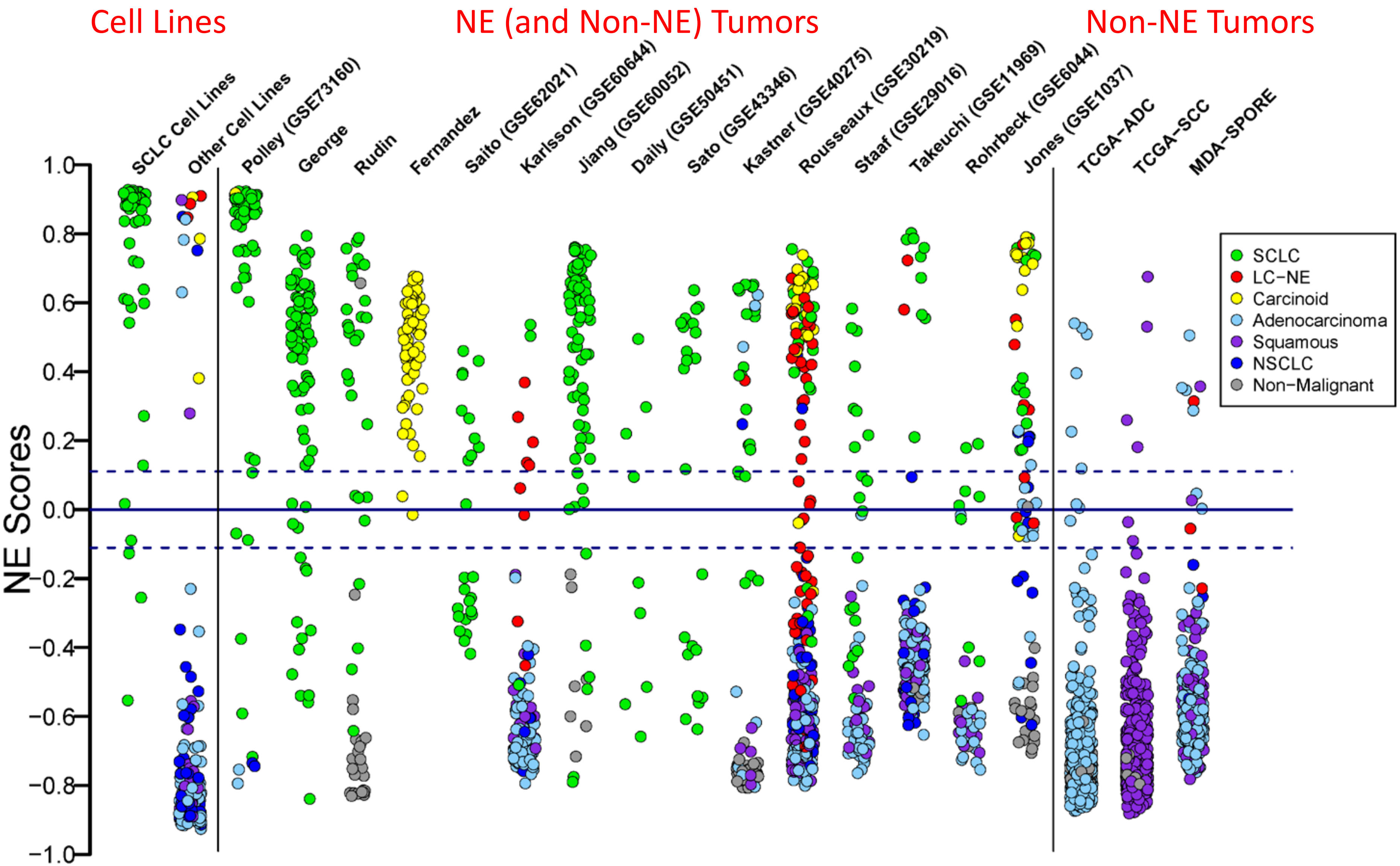
Ranked 52 (excluded from NE signature)

Lung NE Signature Applied to All Cell Lines

Lung NE Score =
(Correl NE – Correl Non-NE) / 2



Lung NE Signature Applied to Lung Primary Tumor Datasets



Conclusions

Cell Line Sequencing

- We performed WES, WGS, and RNAseq on 172 NSCLC and 72 SCLC cell lines as well as 32 PDXs.
- 17 cancer gene mutations were found on average in normal-matched SCLC cell lines
- The SCLC sequencing data have been submitted to cBioPortal

NE Expression Signature

- We developed a gene expression signature distinguishing NE from non-NE lung tumors.
- The signature is highly correlated with four genes used for immunostains: CHGA, INSM1, NCAM1, SYP
- When applied to SCLC tumors, we find a relatively large subset (10-20%) with a low NE score
- These SCLCs appear to correspond to the "variant" subgroup of SCLCs

Ongoing Studies

- Complete WES and RNAseq analyses
- Western blot analyses of key proteins (ASCL1, NEUROD1, MYC family members and Myc partners, RB)
- Mass spectroscopy proteomic analyses of SCLC (Yonghao Yu, UTSW)
- RNAseq analyses of SCLC lines grown as xenografts for comparison to tissue culture
- Comparison of SCLC cell lines and xenografts to SCLC PDXs, CDXs

Acknowledgements

UT Southwestern Medical Center

- John D. Minna
- Adi F. Gazdar
- Kenneth Huffman
- Boning Gao
- Brenda Timmons
- Hans Ghayee
- Yang Xie

British Columbia Cancer Research Centre

- Stephen Lam
- Yuzhuo Wang

Baylor College of Medicine

- David Wheeler
- Caleb Davis
- Kyle Covington
- Chase Miller
- Xiaoyun Liao

NICHD, NIH

- Karel Pacak