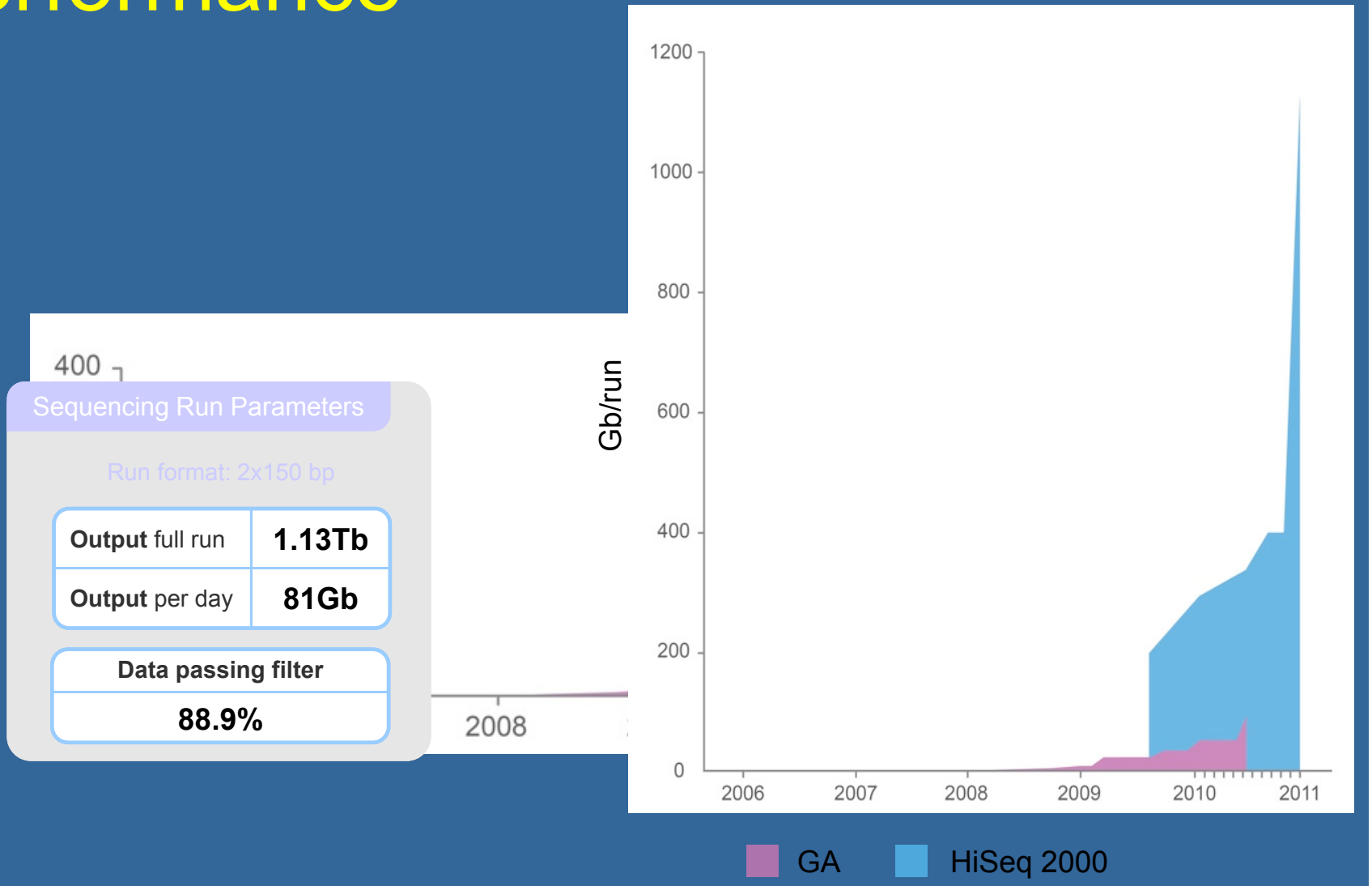
Integration of multiple datasets in head and neck cancer

Analysis of transcriptome and exome sequencing with methylation array data

David I Smith, Ph.D.
Professor
Chairman of the Technology Assessment Group
Center for Individualized Medicine
Mayo Clinic

HiSeq 2000

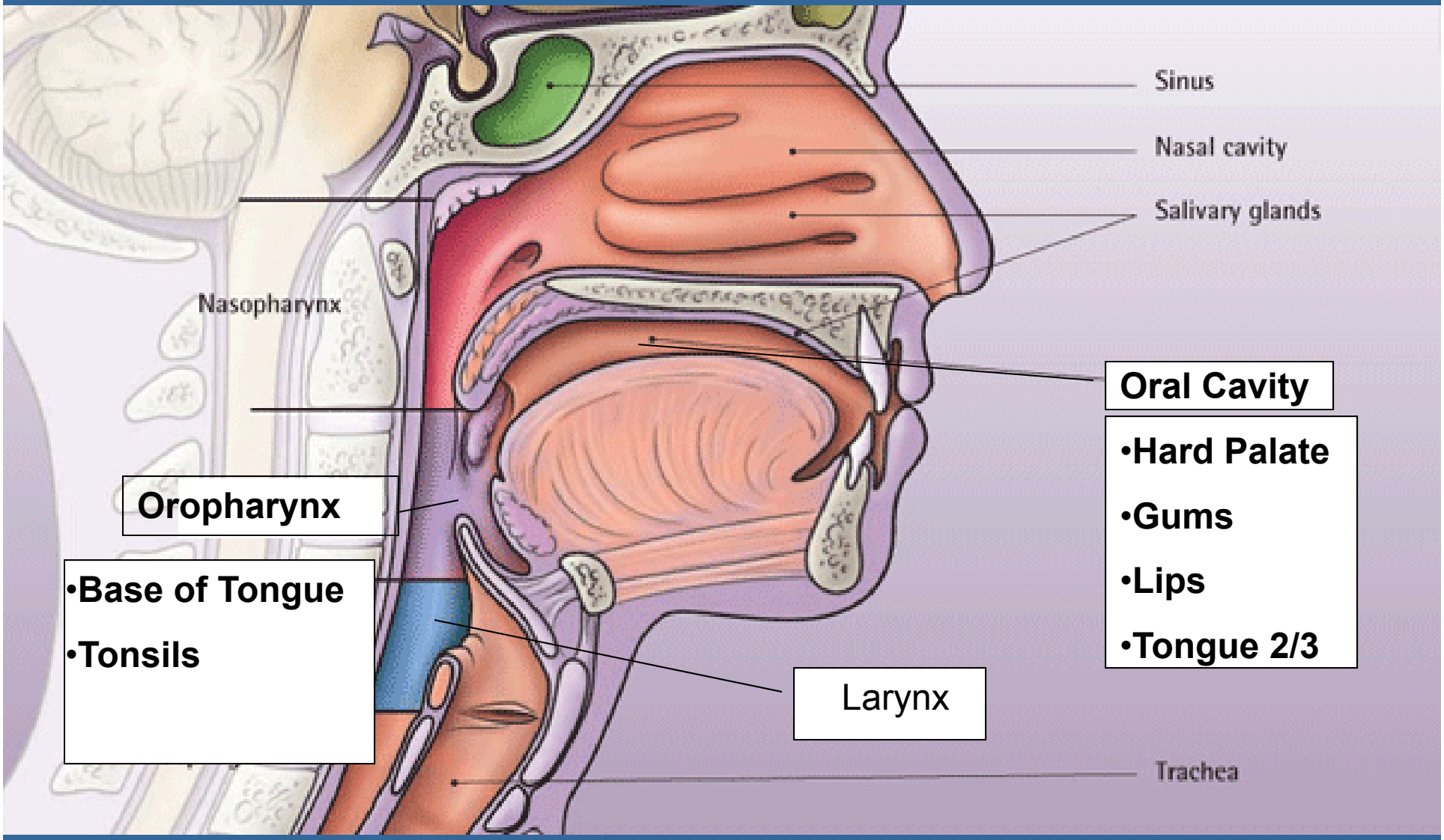
Evolution of Instrument Performance



Uses of Next Generation Sequencing

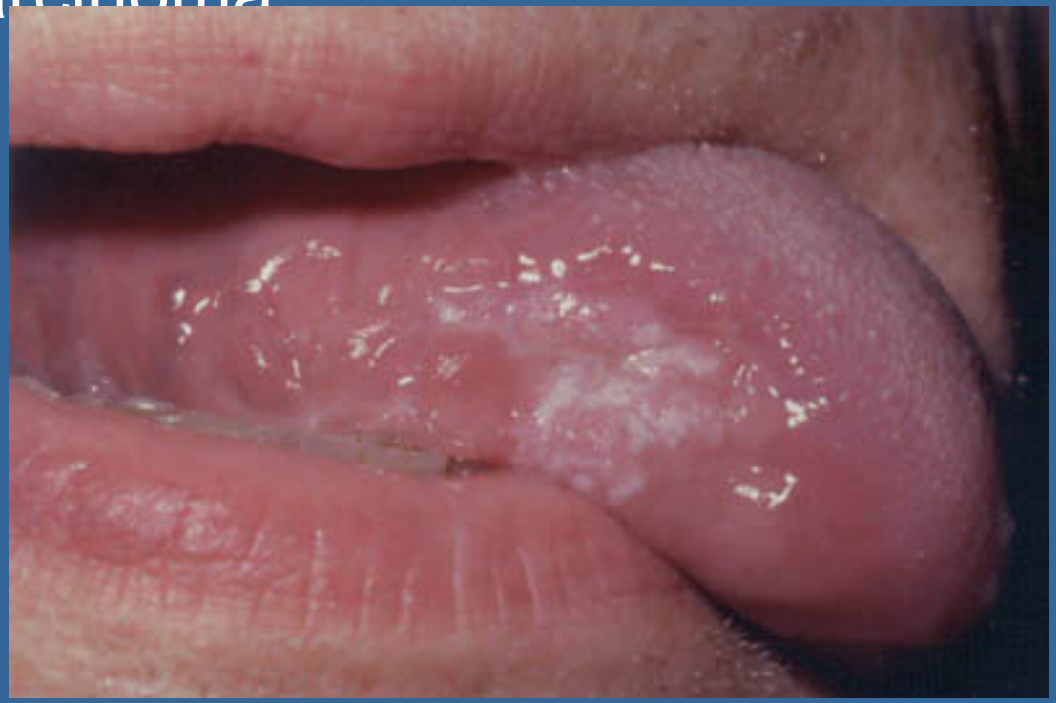
- Whole genome sequencing- requires 100+ gigabases of total sequencing (probably more for cancer genomes)
- Whole exome sequencing- requires enrichment of the exome and then deep sequencing of the resulting 50 Mbs
- Transcriptome sequencing- what is actively being transcribed in the cells of interest
- Methlyation sequencing- characterizing the epigenome, or utilize the Illumina 450K methylation arrays

Head and Neck Cancer



Head and Neck cancer

- 6th most common cancer
- Epithelial Cancers (Carcinomas)
  - Squamous Cell Carcinoma
  - Adenocarcinoma
  - Thyroid
- Mesenchymal
  - (Sarcomas)



Risk Factors

- Smoking and drinking
  - 6-7th decade of life, prolonged exposure
  - All sites

- HPV (16,18)
  - Oropharyngeal SCC (30-90% of patients)
  - Younger patients (<50 years)
  - Lack traditional risk factors
  - Chemo/Radiation Sensitive
  - Lacking strong evidence of consistent viral integration

Head and Neck Function

- Speech/Communication
- Swallowing
- Smell
- Taste
- Vision
- Cosmesis
- Hearing

Treatment

- Overall mortality ~ 60%
  - Depends on Patient, Tumor, and Treatment
- Treatment modalities
  - Surgery: Primary vs Salvage
  - Non operative therapy
    - Radiation: Primary vs adjunct vs salvage
    - Chemotherapy: adjunctive
      - Neoadjuvant (induction), adjuvant (concomitant)

Treatment modality

- Surgery: $30,000-50,000
- Radiation: $100,000-300,000
  - Outcomes roughly equivalent
- Chemotherapy: $100,000-500,000
  - 4% survival benefit in select patients
  - 4% of patients die from treatment
  - Financial incentive to treat with Radiation/Chemo?

HPV not just in the cervix/cervical cancer

- In cervical cancer HPV long-term infection coupled with eventual integration appears to be the trigger causing cervical cancer
- HPV also found in some head and neck cancers, anal cancer, vulva and vaginal cancer, and cancer of the penis
- Role of HPV in these other sites is less well understood, however

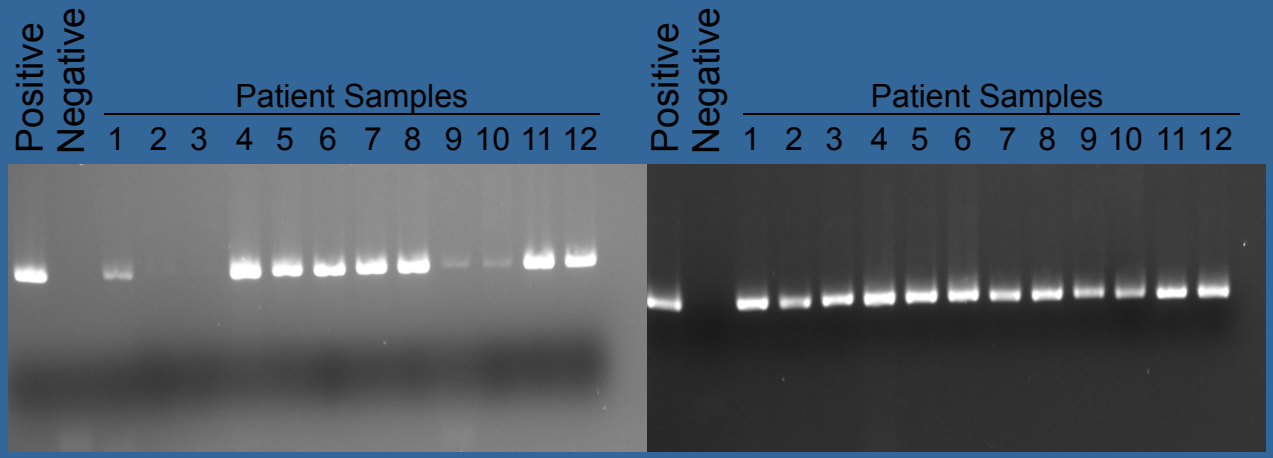
HPV oncogenes

- E6 and E7 genes are consistently expressed in HPV infected cells
- HPV E6 decreases p53 expression
- HPV E7 gene leads to loss of pRB cell cycle arrest
- pRb negatively feeds back on p16
- When pRb is down regulated by HPV E7, p16 is over expressed

Detecting HPV in head and neck cancer

- In situ hybridization for HPV sequences
- PCR for HPV DNA
- Immunohistochemistry for p16 expression (which is elevated in HPV+ cancers)
- Detection of transcriptionally active infections by measuring RNA transcripts, or protein for the oncoproteins E6 and E7

HPV Detection by PCR



Universal Primers **HPV 16 Specific Primers**

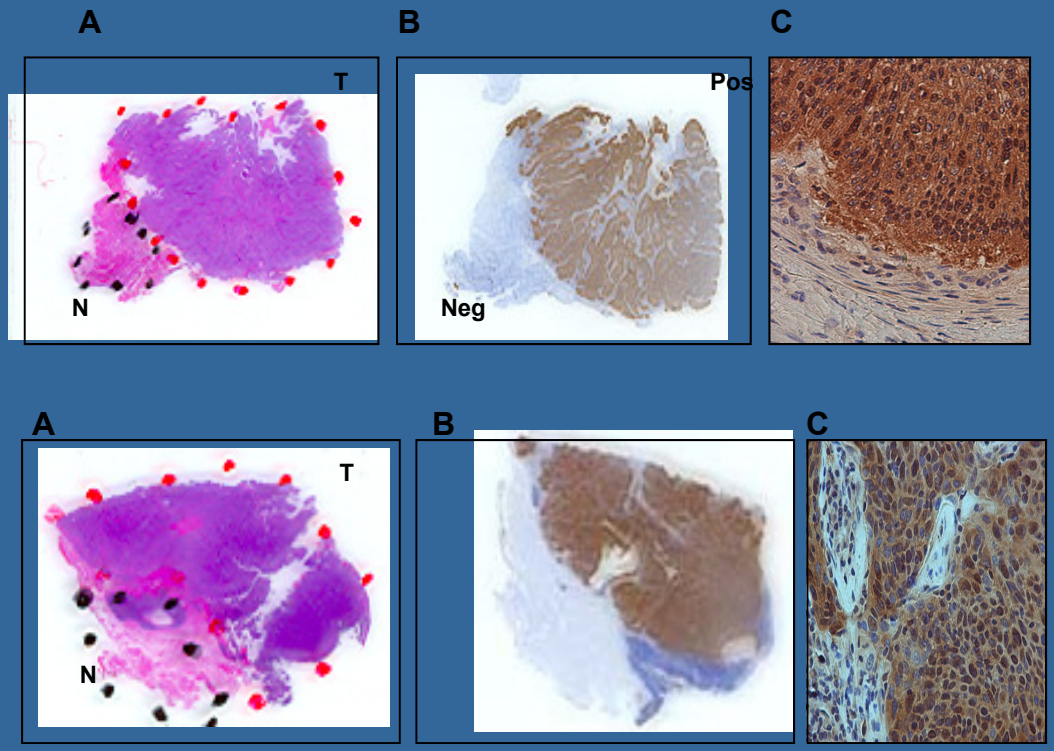
- More Consistent results from Fresh Frozen Tissues versus Paraffin Embedded Samples

HPV Detection by IHC

A: H&E Stain

B: IHC p16^{INK4A}

C: IHC p16INK4A (40X)



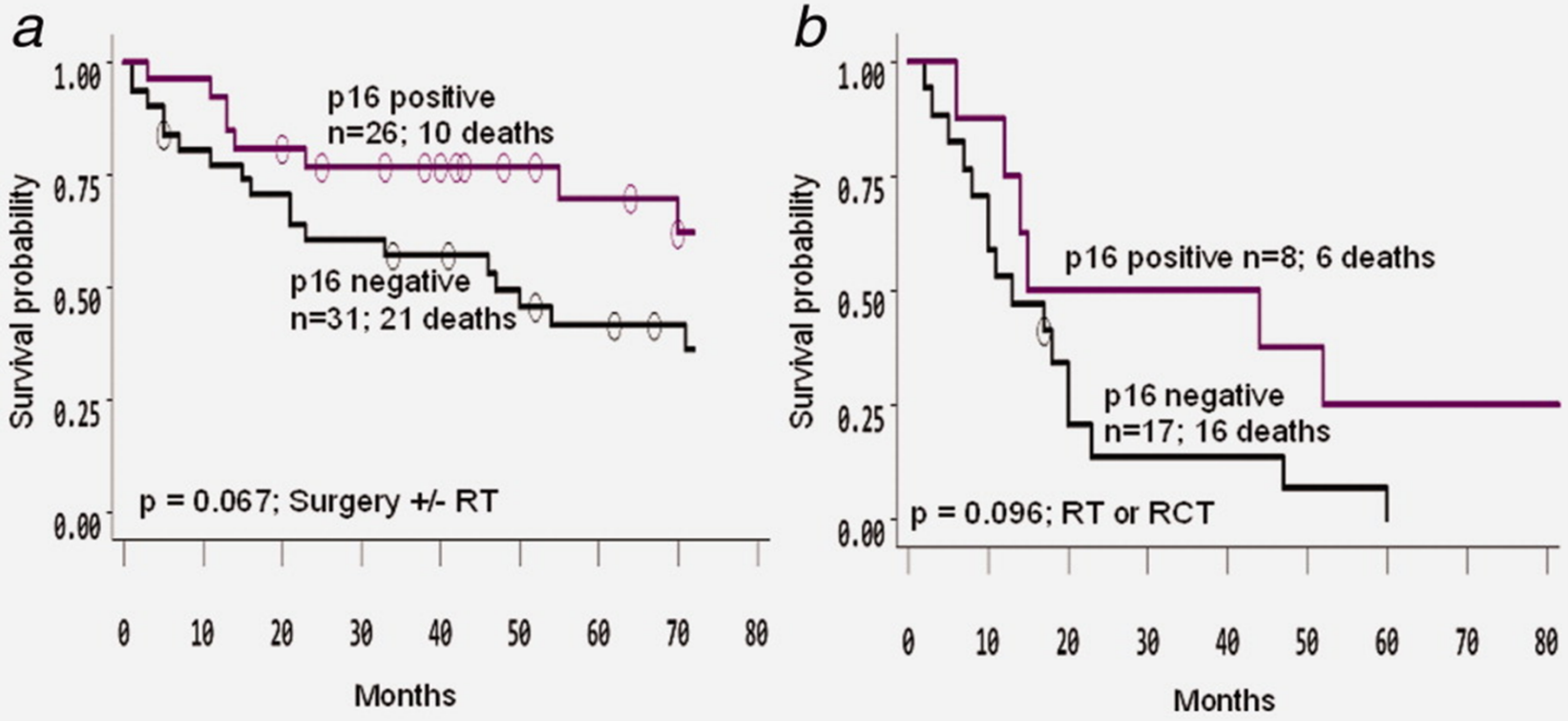
HPV E6 and E7 Expression by Q-RT PCR

- Not routinely done
- Difficulty in doing this out of paraffin
- New technologies to do this- Nanostring, Fluidigm, or Advanced Cell Diagnostics (RNAscope)

HPV Infection Rates

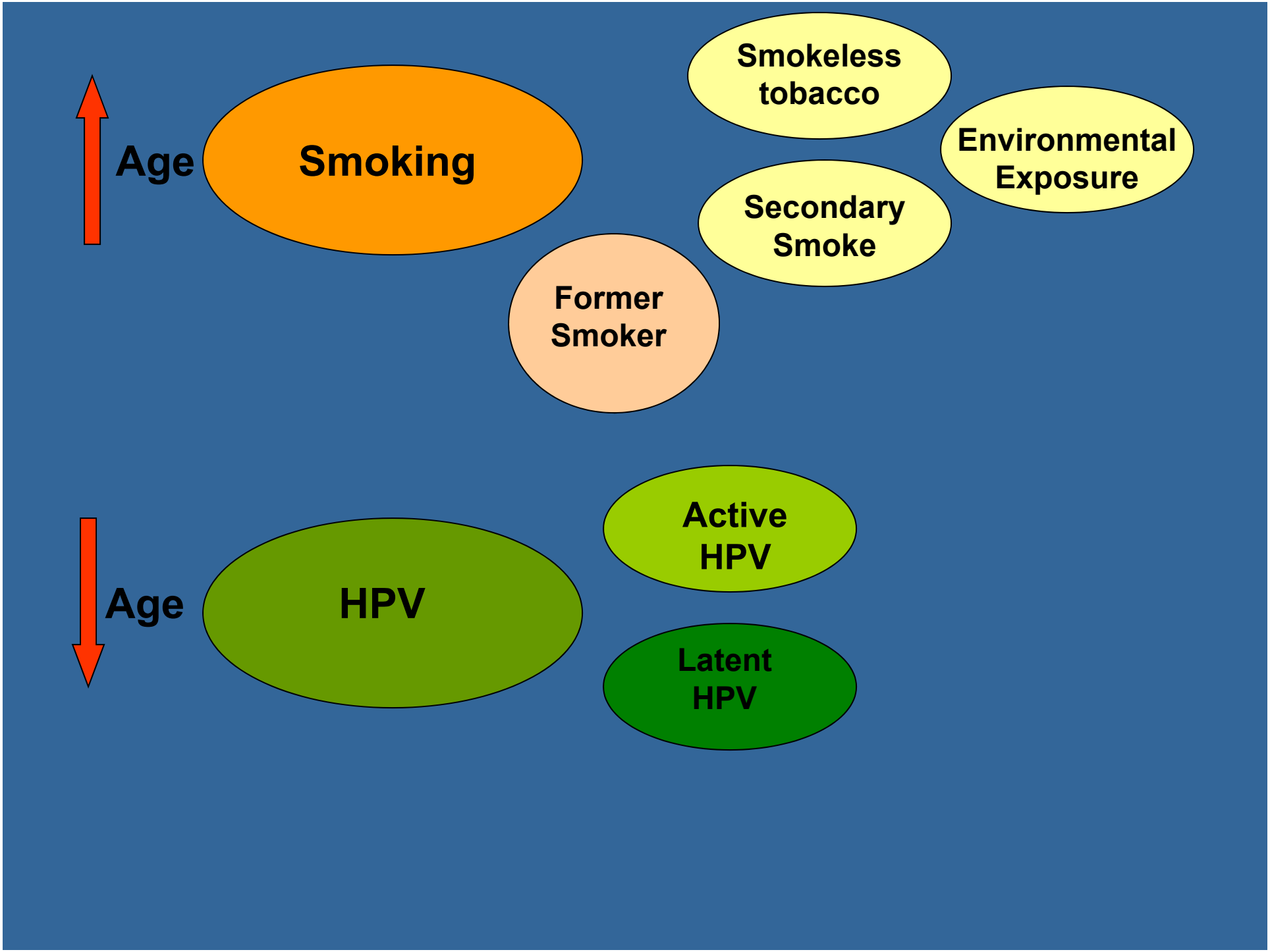
- High rates of HPV positivity in BTT (oropharyngeal) tumors (26-91%), Is this because HPV rates differ in different populations or is this a technical problem?
- Low rates of HPV positivity in Oral Cavity tumors (3-11%)

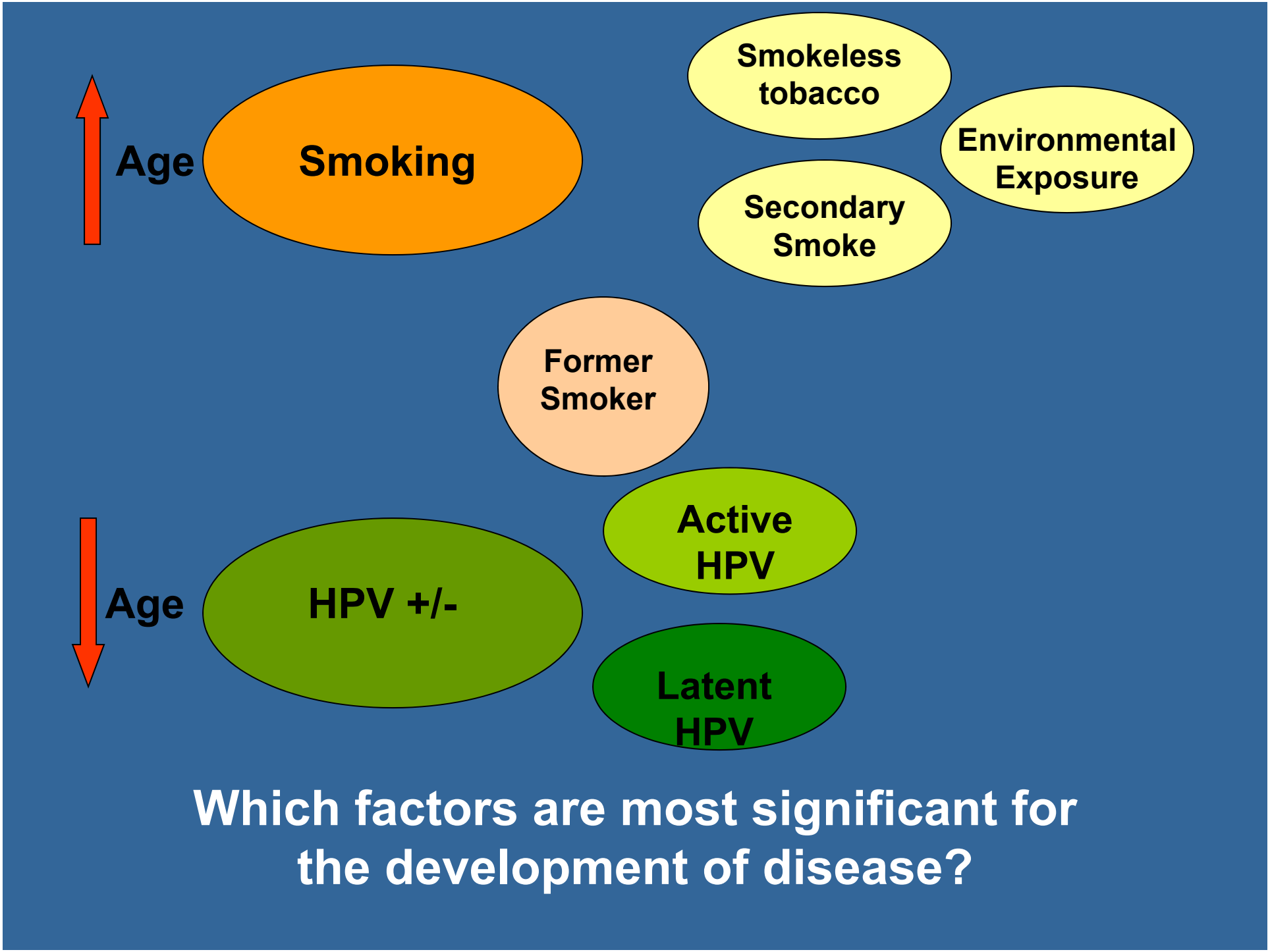
- Oropharyngeal HPV+
  - HPV16,18 most common, other sub-types?


a
1.00
0.75
0.50
0.25
0.00
Survival probability
p16 positive
n=26; 10 deaths
p16 negative
n=31; 21 deaths
p = 0.067; Surgery +/- RT
0
10
20
30
40
50
60
70
80
Months
b
1.00
0.75
0.50
0.25
0.00
Survival probability
p16 positive n=8; 6 deaths
p16 negative
n=17; 16 deaths
p = 0.096; RT or RCT
0
10
20
30
40
50
60
70
80
Months

Two distinct entities of oropharyngeal cancer?

- HPV minus – generally older smokers and drinkers. Much worse clinical outcome due to more DNA damage
- HPV positive- contains those individuals who are both younger and many are non-drinkers/non-smokers. Less DNA damage, hence possibly more treatable?
- Not clean groupings as there are smokers/drinkers that are also HPV positive


Age
Smoking
Smokeless tobacco
Environmental Exposure
Secondary Smoke
Former Smoker
Age
HPV
Active HPV
Latent HPV

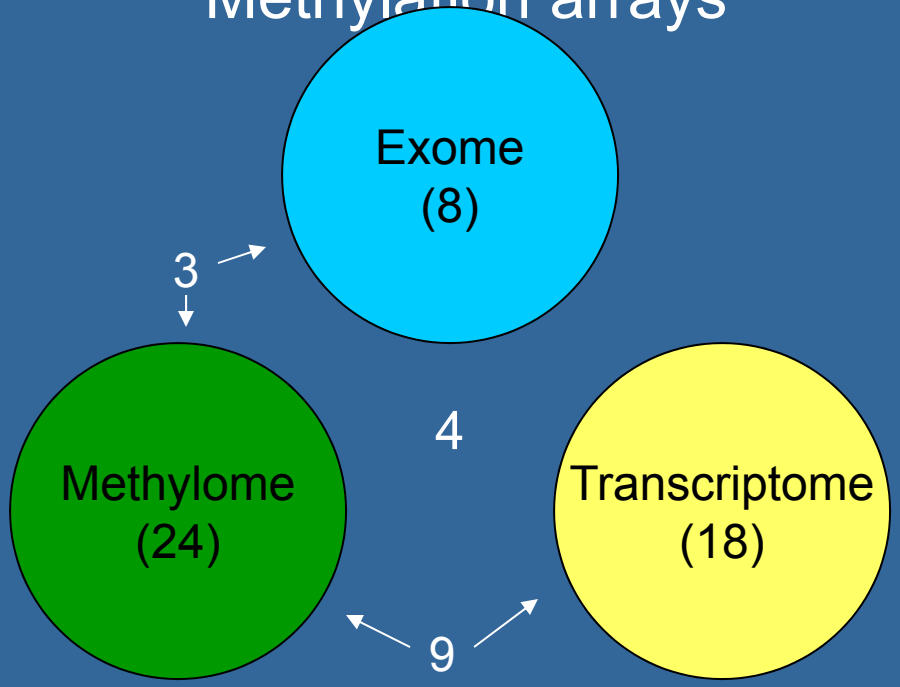

Age
Smoking
Smokeless tobacco
Environmental Exposure
Secondary Smoke
Former Smoker
Active HPV
Age
HPV +/-
Latent HPV
Which factors are most significant for the development of disease?

Application of Massively Parallel DNA sequencing and related technologies

mRNA-Seq / nCounter (or other) validation

Exome sequencing

Methylation arrays



Application of Massively Parallel DNA sequencing and related technologies

mRNA-Seq / nCounter validation

Exome sequencing

Methylation arrays

Clinical Application - Tool Development

Patient knowledge of technologies - expectations

Discovery - Experimental

Clinical Applications

- Supplement current technologies for diagnosis (detection of HPV)
- Develop markers for patient stratification based on most significant risk factors
- Develop tools for early detection of primary disease and recurrence

Experimental Interests

- Transcript dysregulation
  - Risk factors, sites, disease behavior
- Non-coding transcripts
- Fusion transcripts
- Methylation and mutations – define mechanisms of genome dysregulation

mRNA-Seq Transcriptional profiling of Oropharyngeal SCC

- 18 Tumor and Patient Matched Normal Tissue Samples
- Illumina GAIIx (12 samples) (oligo dT-primed) and SOLiD (6 samples) (Ribominus removal of rRNA species)
- ~65 million reads per sample (200 million for SOLiD RNAseq)
- Data analysis –
  - Genesifter, Geospiza, Seattle WA

RNAseq on T/N Pairs

- Illumina GA IIx or SOLiD 4

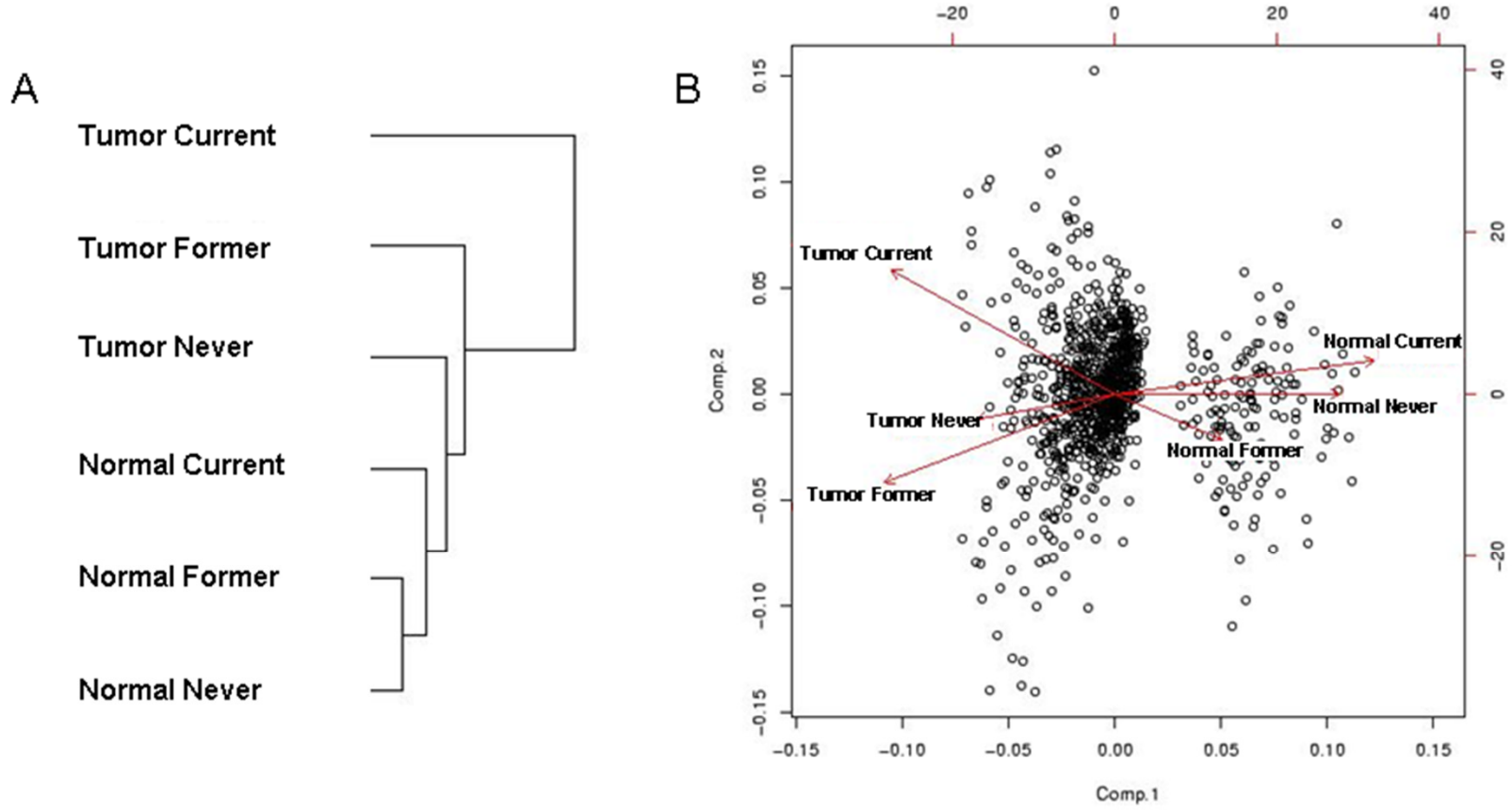
	Age	Gender	Illumina/SOLiD
Current Smoking	56	M	Illumina
	73	M	Illumina
	54	M	Illumina
	65	F	Illumina
	48	M	SOLiD
Former Smoking	48	M	Illumina
	38	M	Illumina
	64	F	Illumina
	66	M	Illumina
	60	F	Illumina
	69	M	SOLiD
	62	M	SOLiD
Never Smoking	46	M	Illumina
	49	F	Illumina
	73	M	Illumina
	48	M	SOLiD
	55	M	SOLiD
	54	F	SOLiD

Develop markers for patient stratification based on most significant risk factors

- Patients present with overlapping risk factors
- Patient self-reporting of tobacco exposure is often unreliable
- Lack of mechanism for role of HPV

Global Gene Expression mRNA-Seq

patients grouped by smoking status

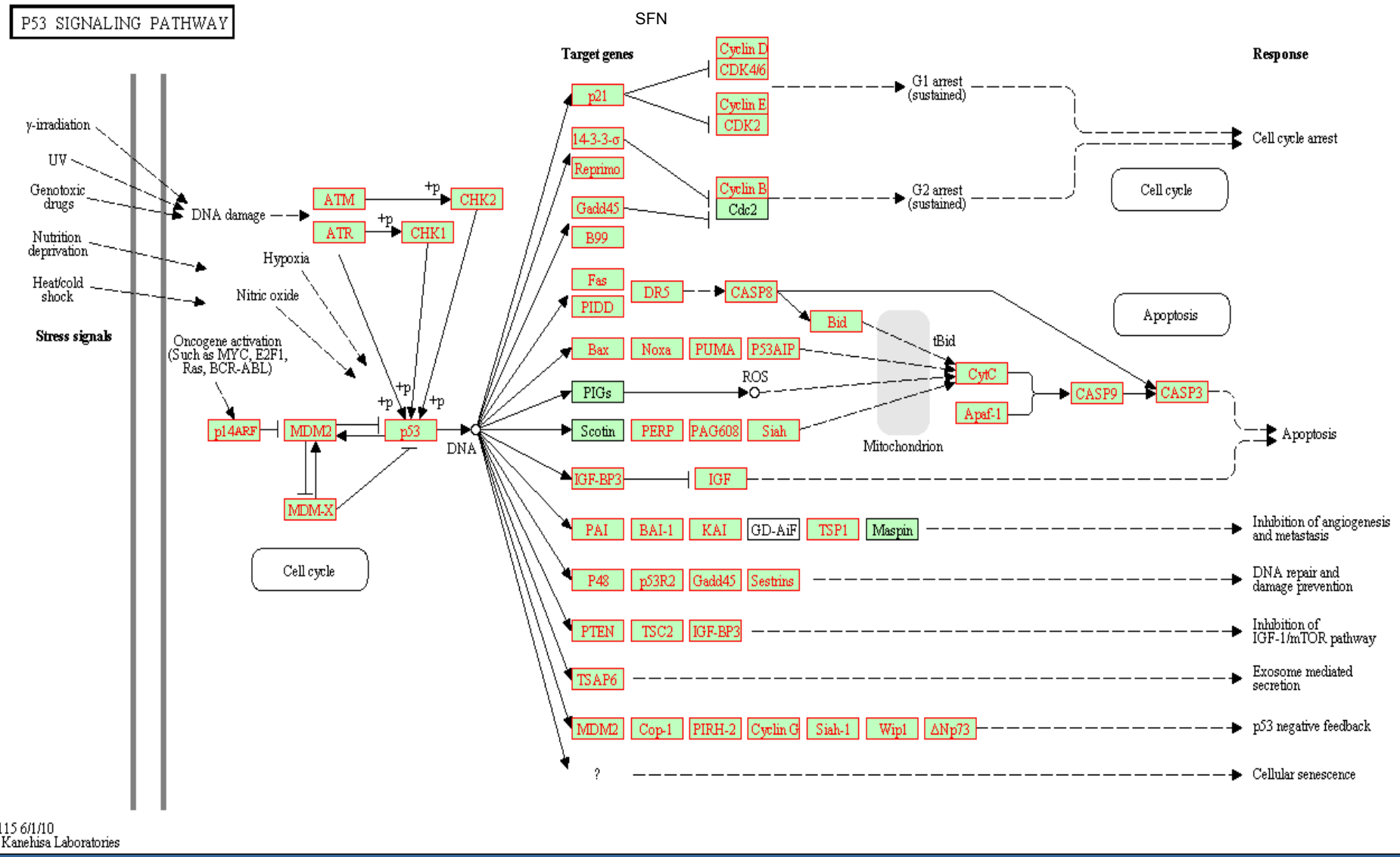


Analysis of transcriptional profiling

- Differentially regulated genes
  - HOX, MMPs, T cell signaling, Immune responsive gene targets
- Pathway analysis (Genesifter)
  - p53 signaling pathway
  - Is there a difference between groups divided by risk factor

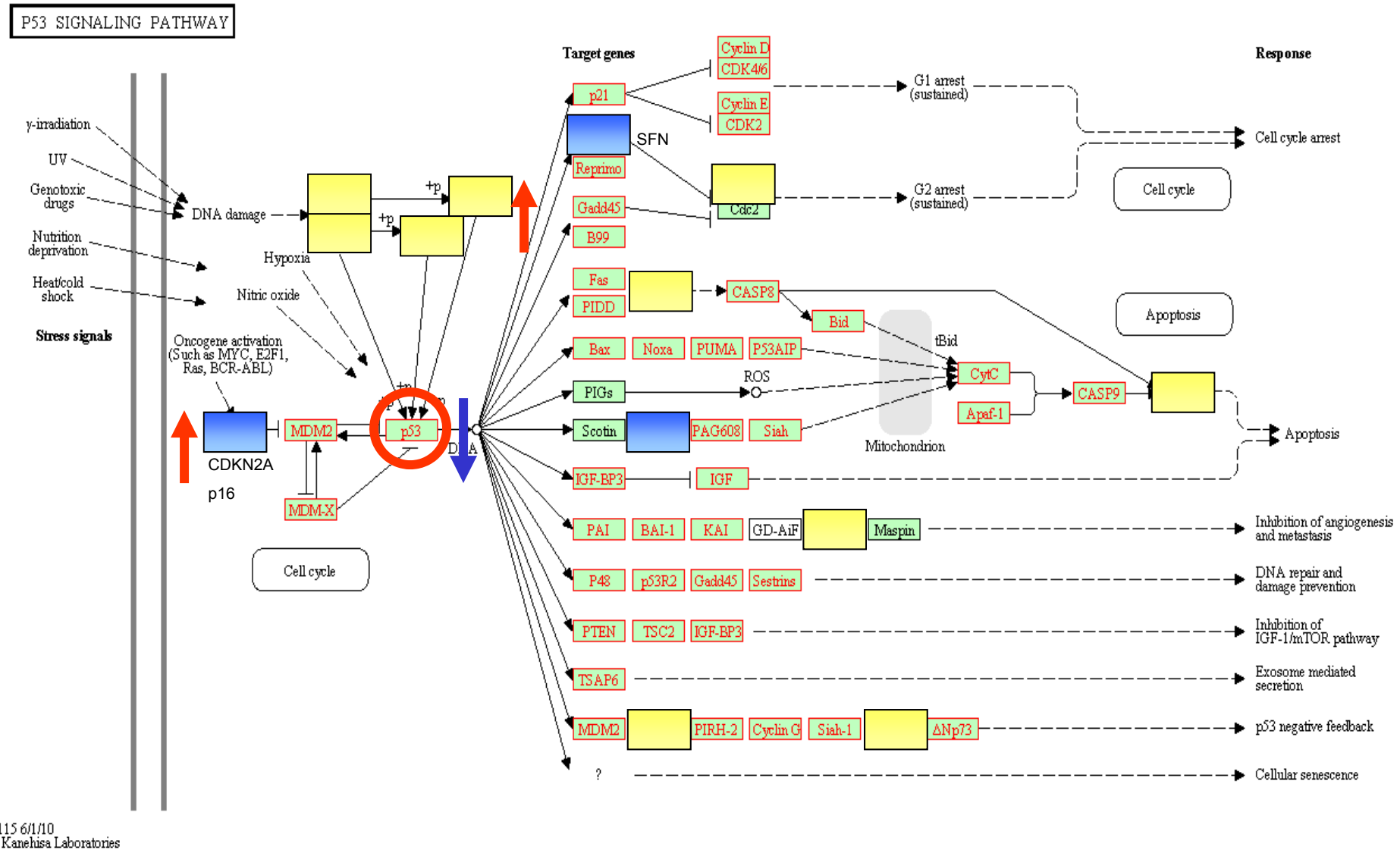
Regulatory Non-coding Transcripts

Transcriptome Analysis : Differential T/N

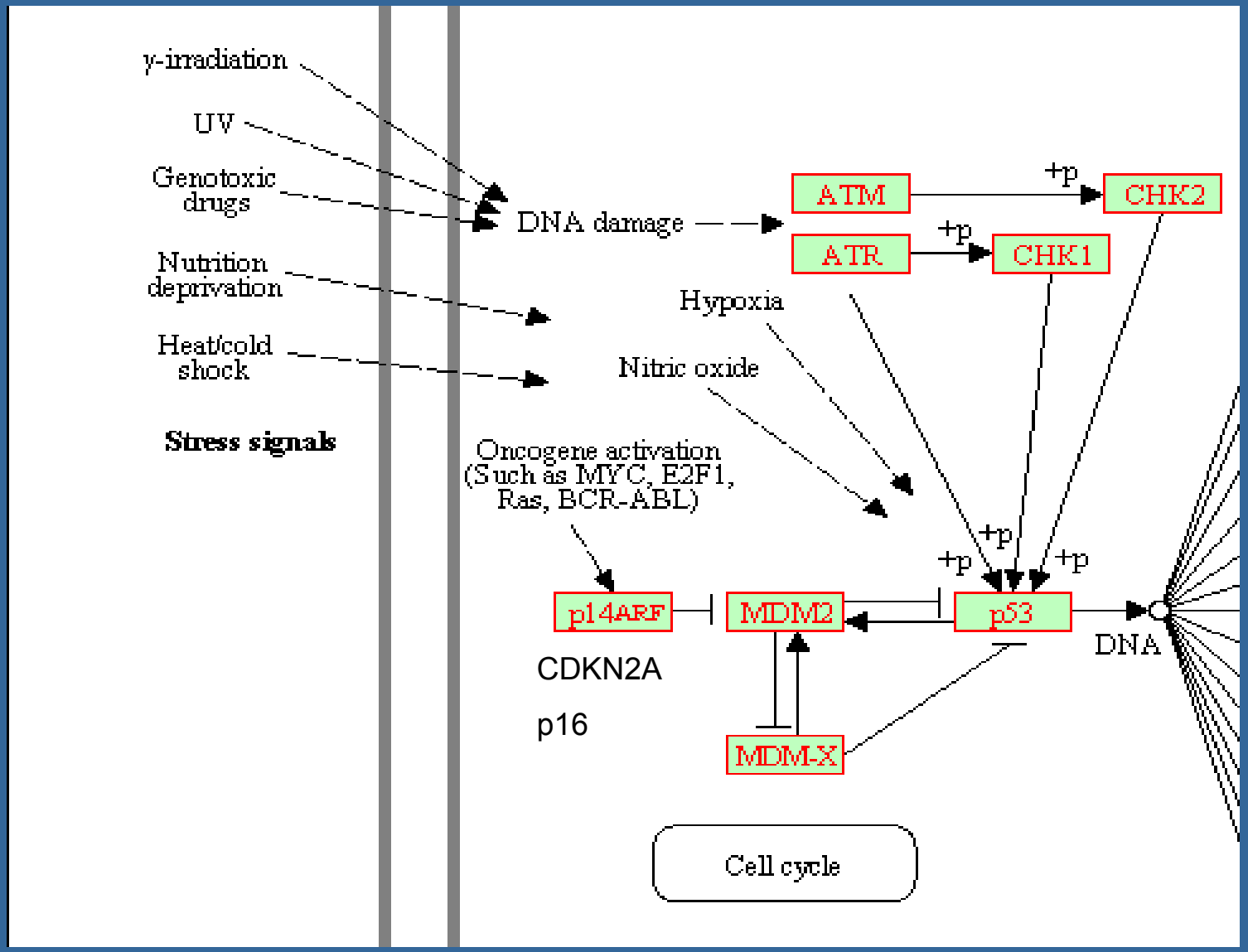


Current Smokers

Never smokers



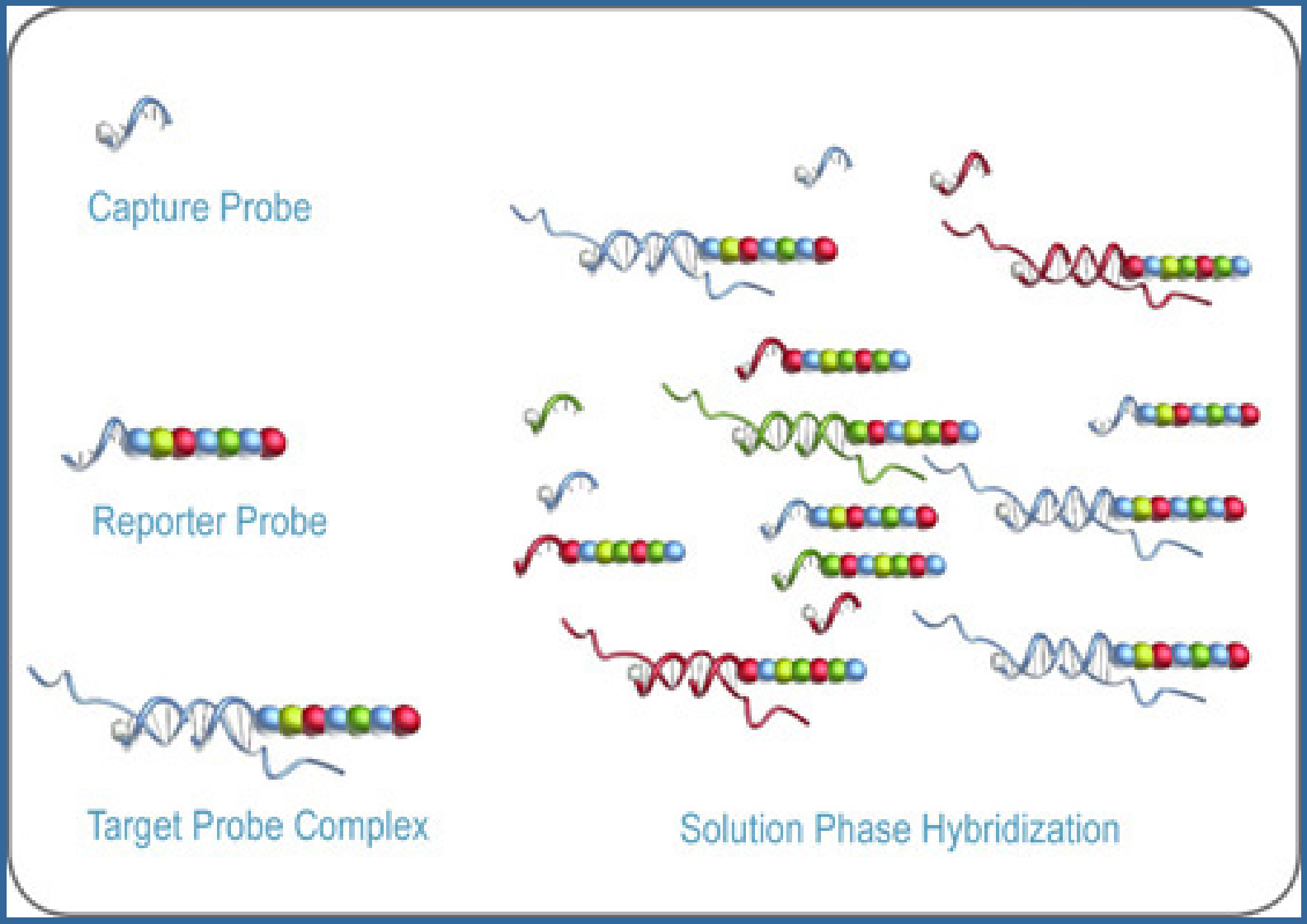
p53 Signaling

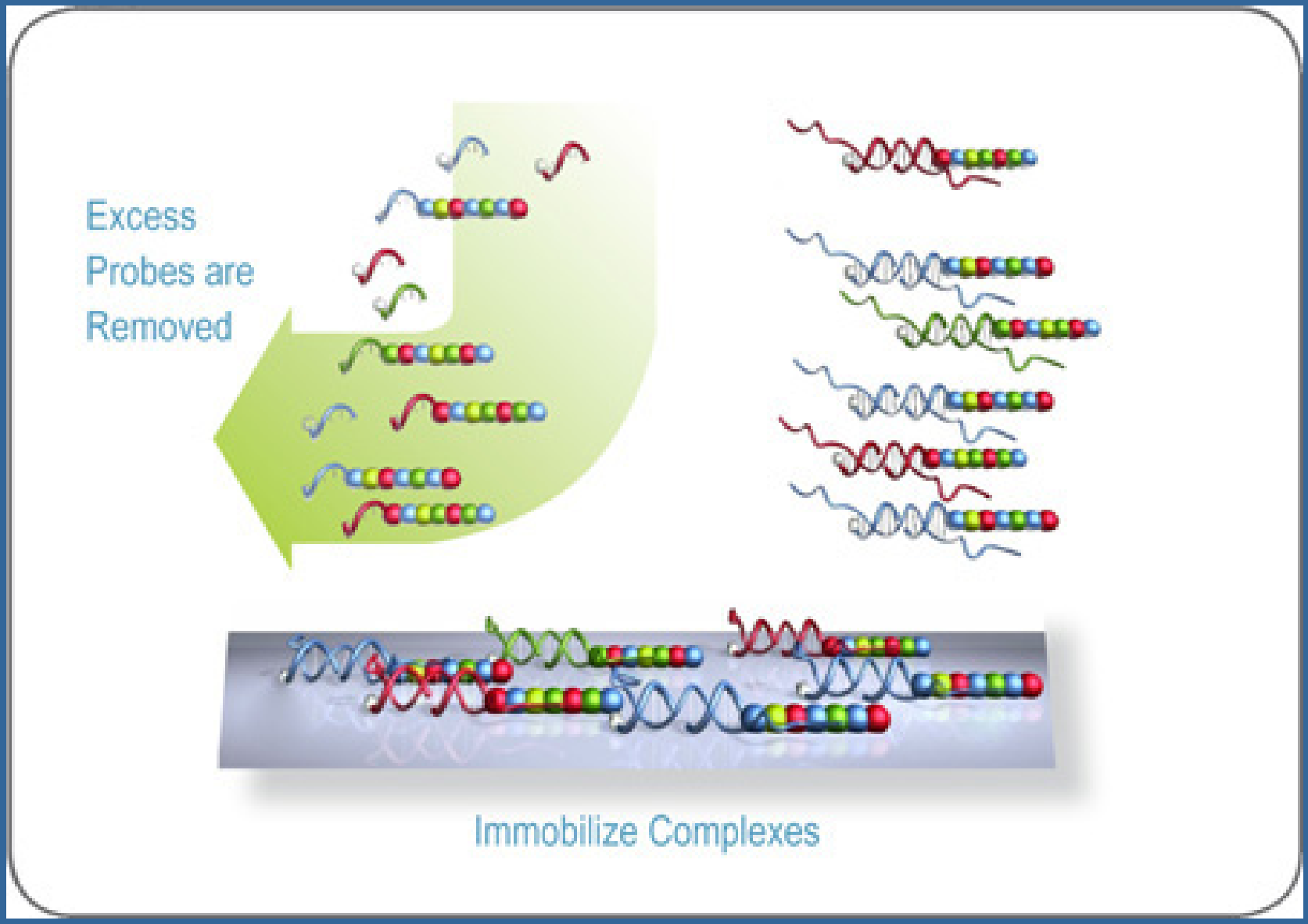


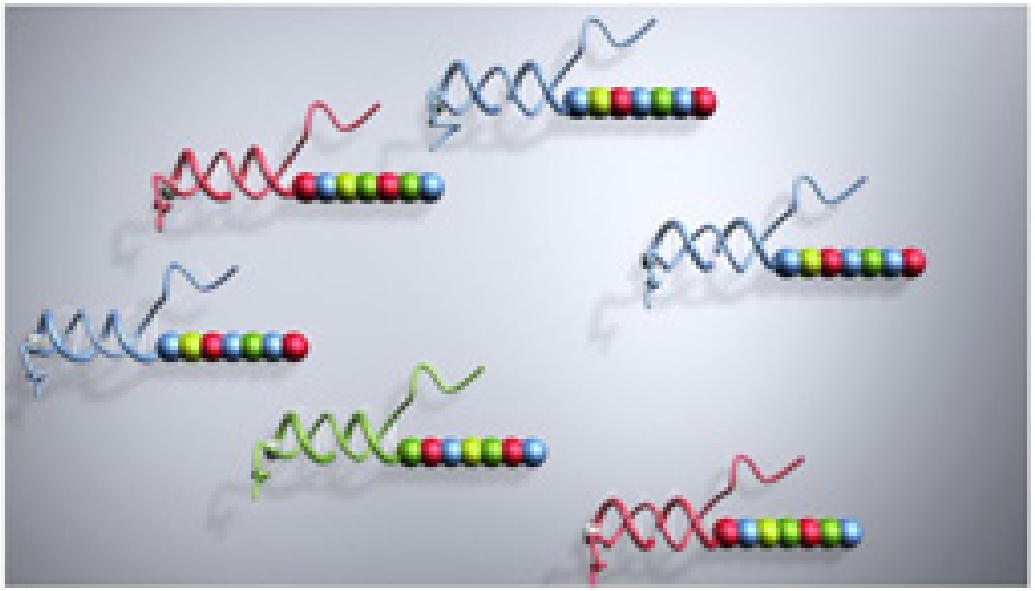
Validation Experiments

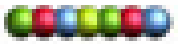
- Small sample size restricts analysis
- Validate in larger number of tumor/normal pairs
- Nanostring nCounter digital counter
  - 96 gene codesets
  - 44 sample pairs
  - 100ng input RNA
  - Compare fresh frozen to FFPE

nCounter Codesets




Excess
Probes are
Removed
Immobilize Complexes



Color Code	Counts	Identity
	3	XLSA
	2	FOX5
	1	INSULIN

nCounter HPV gene target analysis

- Evaluate expression patterns of genes with implicated correlation to HPV infection in head and neck cancer
- Group patients based on HPV16- E7 transcriptional levels

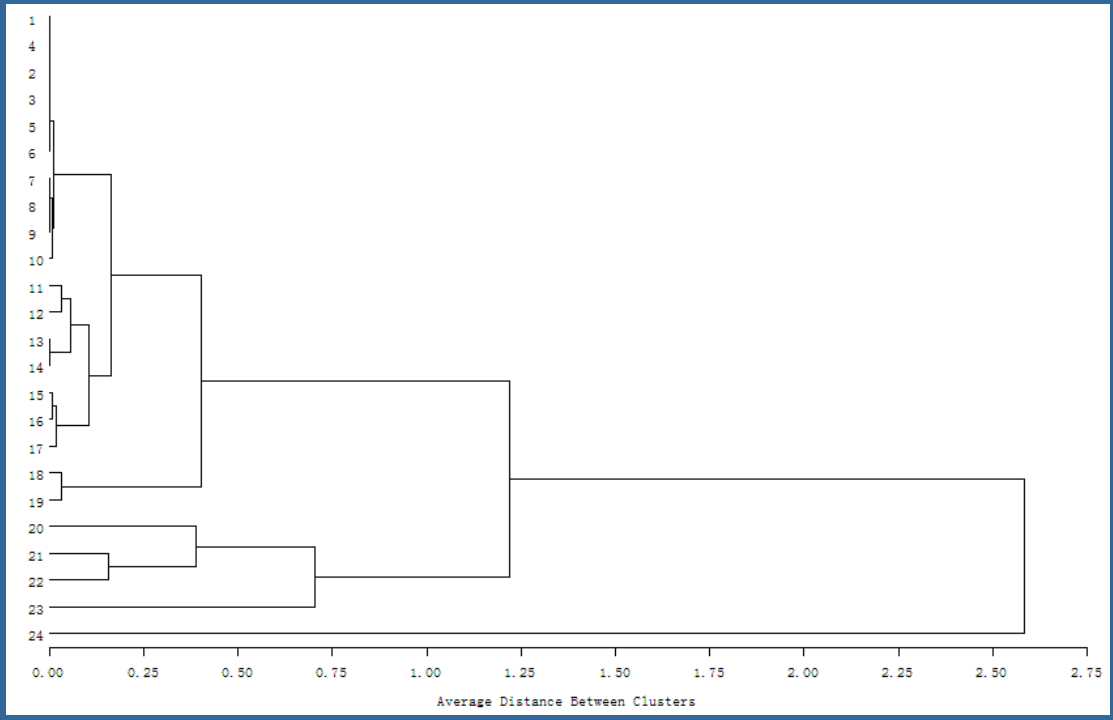
Average linkage cluster analysis HPV16- E7 by qPCR

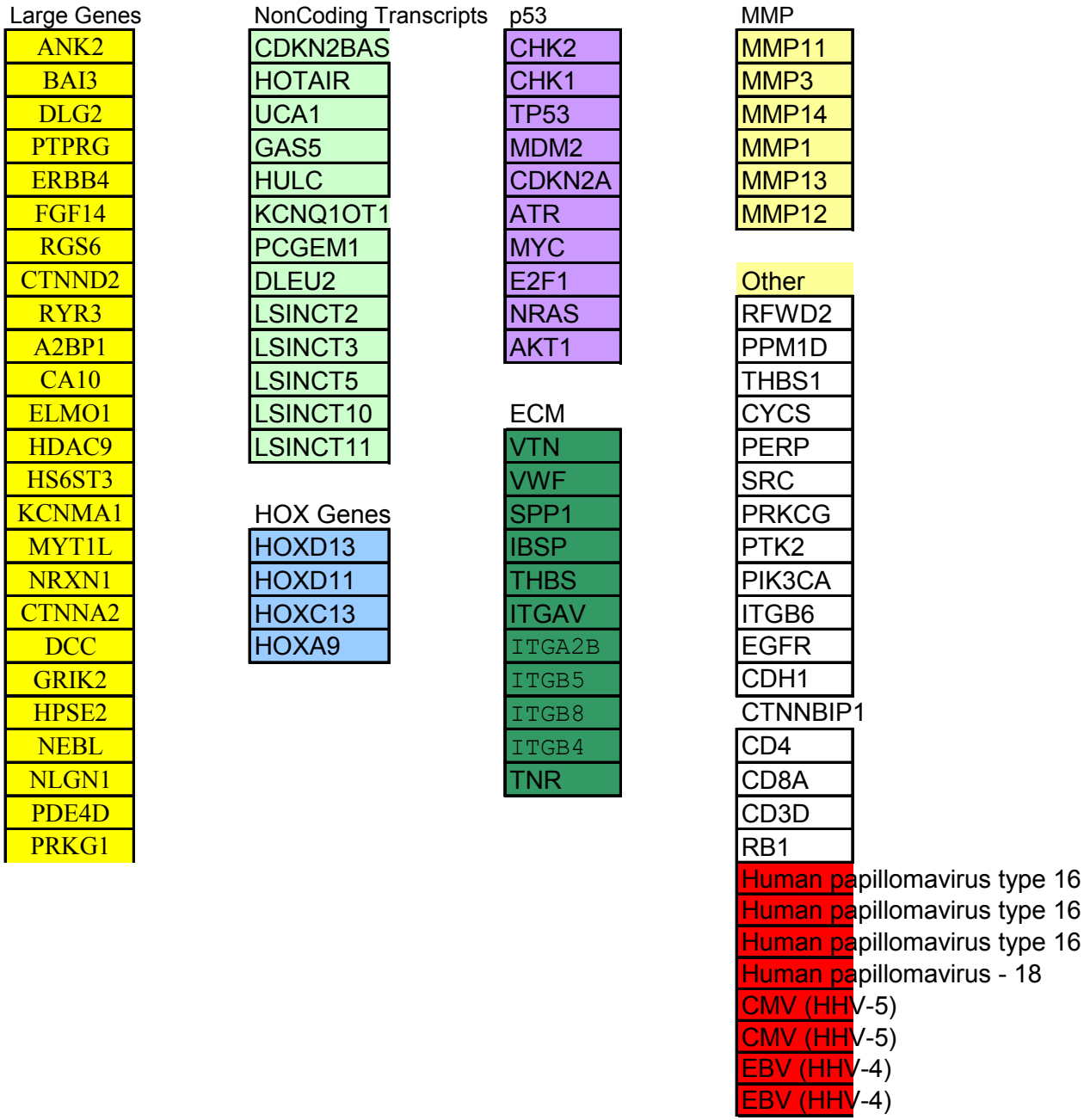
HPV status grouping	Patient	HPV16 E6	HPV16 E7	Subsite	Age	Smoking Status	T stage	N Stage	M Stage	Grade
Negative / Latent	1	2.21E-06	2.05E-06	Tonsil	73	Never	2	2b	0	3
	4	2.56E-05	3.17E-05	BOT	59	Former	3	2b	0	3
	2	0.000136	0.00102	Tonsil	60	Never	1	2b	0	3
	3	0.000162	0.001197	Tonsil	65	Former	1	1	0	3
	5	0.00062	0.001232	Tonsil	46	Current	1	2b	0	NR
	6	0.001333	0.002012	Tonsil	45	Never	4a	2b	1	3
	7	0.004072	0.005839	Tonsil	47	Former	4a	2c	0	4
	8	0.004714	0.00821	Tonsil	73	Former	3	2b	0	3
	9	0.004824	0.032355	BOT	62	Former	2	2b	0	3
	10	0.008126	0.019382	Tonsil	61	Never	1	2b	0	3
Low	11	0.041836	0.040049	Tonsil	56	Former	2	2b	0	3
	12	0.059295	0.102554	BOT	78	Never	3	2b	0	4
	13	0.079654	0.123738	BOT	75	Former	2	2b	0	3
	14	0.080307	0.153541	BOT	62	Never	1	2a	0	3
	15	0.116239	0.211119	Tonsil	63	Former	3	2b	0	4
	16	0.12027	0.267703	Tonsil	48	Never	1	0	0	3
	17	0.127278	0.17129	Tonsil	49	Current	3	2b	0	3
	18	0.247945	0.263652	BOT	39	Never	2	2a	0	3
	19	0.26658	0.335911	Tonsil	63	Never	2	2b	0	3
High	20	0.475037	0.470172	BOT	67	Current	2	2b	0	3
	21	0.644739	0.547167	BOT	55	NR	1	2a	0	3
	22	0.731662	1.472993	BOT	57	Never	2	0	0	3
	23	1.593043	1.774166	BOT	62	Former	2	2c	0	3
	24	0.995721	2.126597	BOT	57	Current	2	2c	0	4

Negative / Latent E7

Low E7

High E7

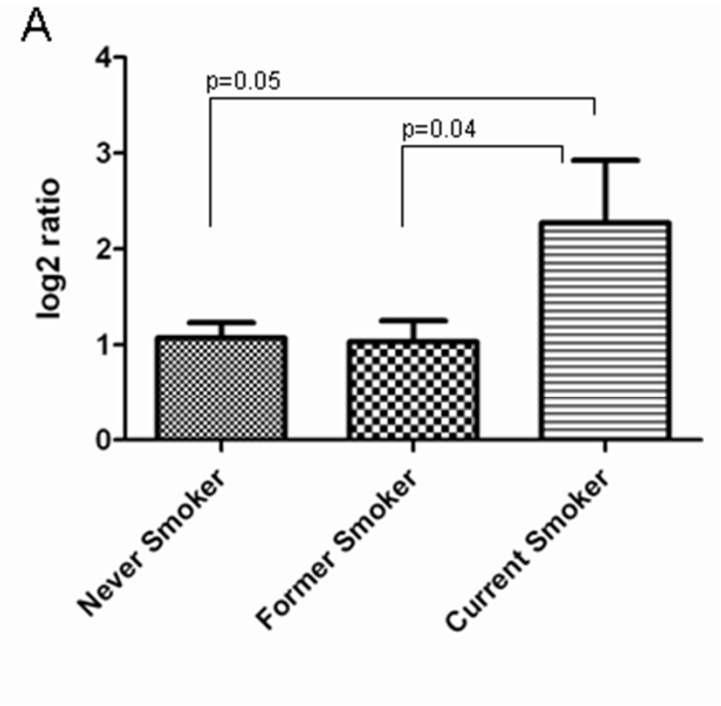


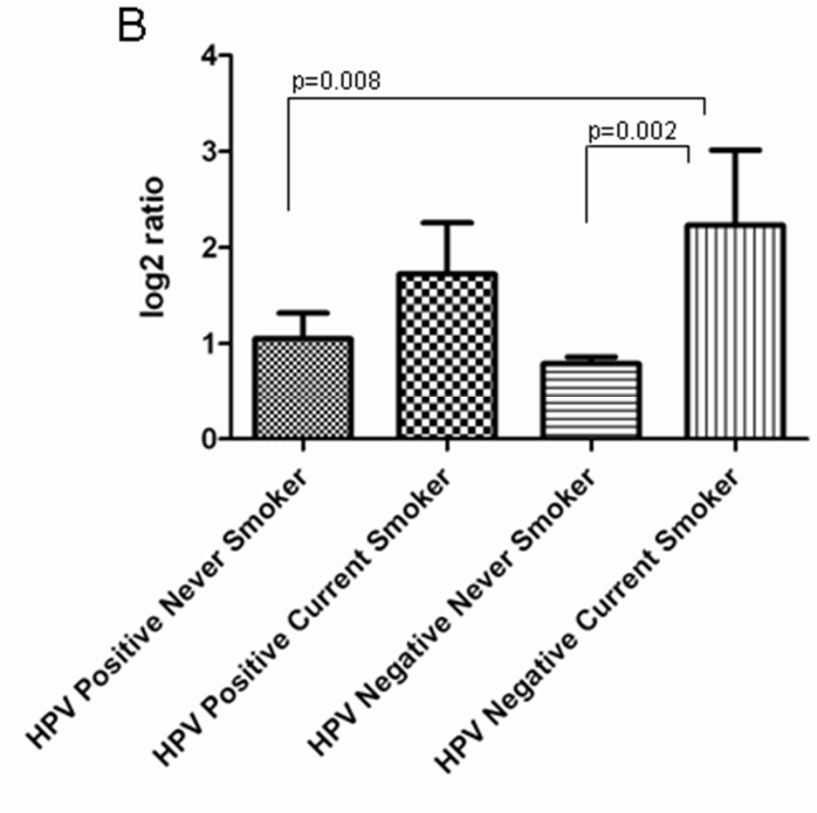

Large Genes
ANK2
BAI3
DLG2
PTPRG
ERBB4
FGF14
RGS6
CTNND2
RYR3
A2BP1
CA10
ELMO1
HDAC9
HS6ST3
KCNMA1
MYT1L
NRXN1
CTNNA2
DCC
GRIK2
HPSE2
NEBL
NLGN1
PDE4D
PRKG1
NonCoding Transcripts
CDKN2BAS
HOTAIR
UCA1
GAS5
HULC
KCNQ1OT1
PCGEM1
DLEU2
LSINCT2
LSINCT3
LSINCT5
LSINCT10
LSINCT11
HOX Genes
HOXD13
HOXD11
HOXC13
HOXA9
p53
CHK2
CHK1
TP53
MDM2
CDKN2A
ATR
MYC
E2F1
NRAS
AKT1
ECM
VTN
VWF
SPP1
IBSP
THBS
ITGAV
ITGA2B
ITGB5
ITGB8
ITGB4
TNR
MMP
MMP11
MMP3
MMP14
MMP1
MMP13
MMP12
Other
RFWD2
PPM1D
THBS1
CYCS
PERP
SRC
PRKCG
PTK2
PIK3CA
ITGB6
EGFR
CDH1
CTNNBIP1
CD4
CD8A
CD3D
RB1
Human papillomavirus type 16
Human papillomavirus type 16
Human papillomavirus type 16
Human papillomavirus - 18
CMV (HHV-5)
CMV (HHV-5)
EBV (HHV-4)
EBV (HHV-4)

Results

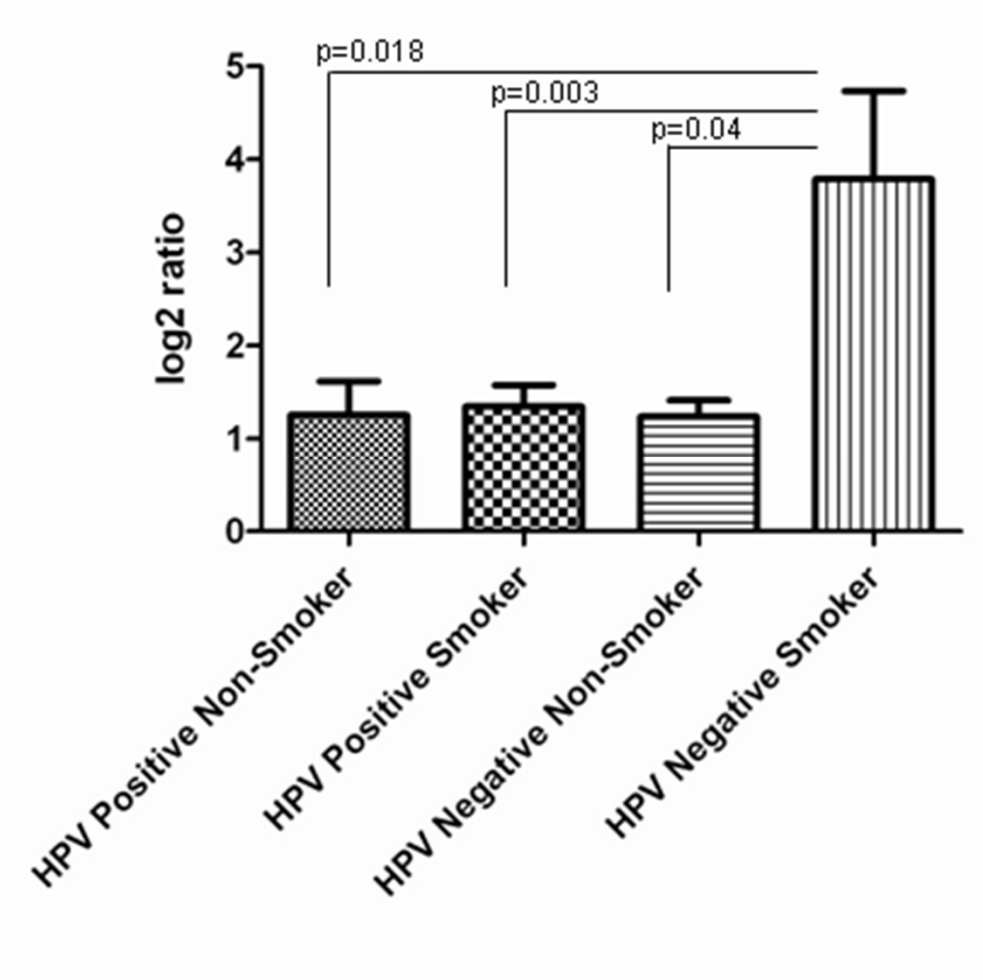
- Did not work out of paraffin at all
- Appears to work for more abundantly expressed transcripts, not too good for low abundance
- Worked great for HPV E6 and E7

ATR ataxia telangiectasia and Rad3 related

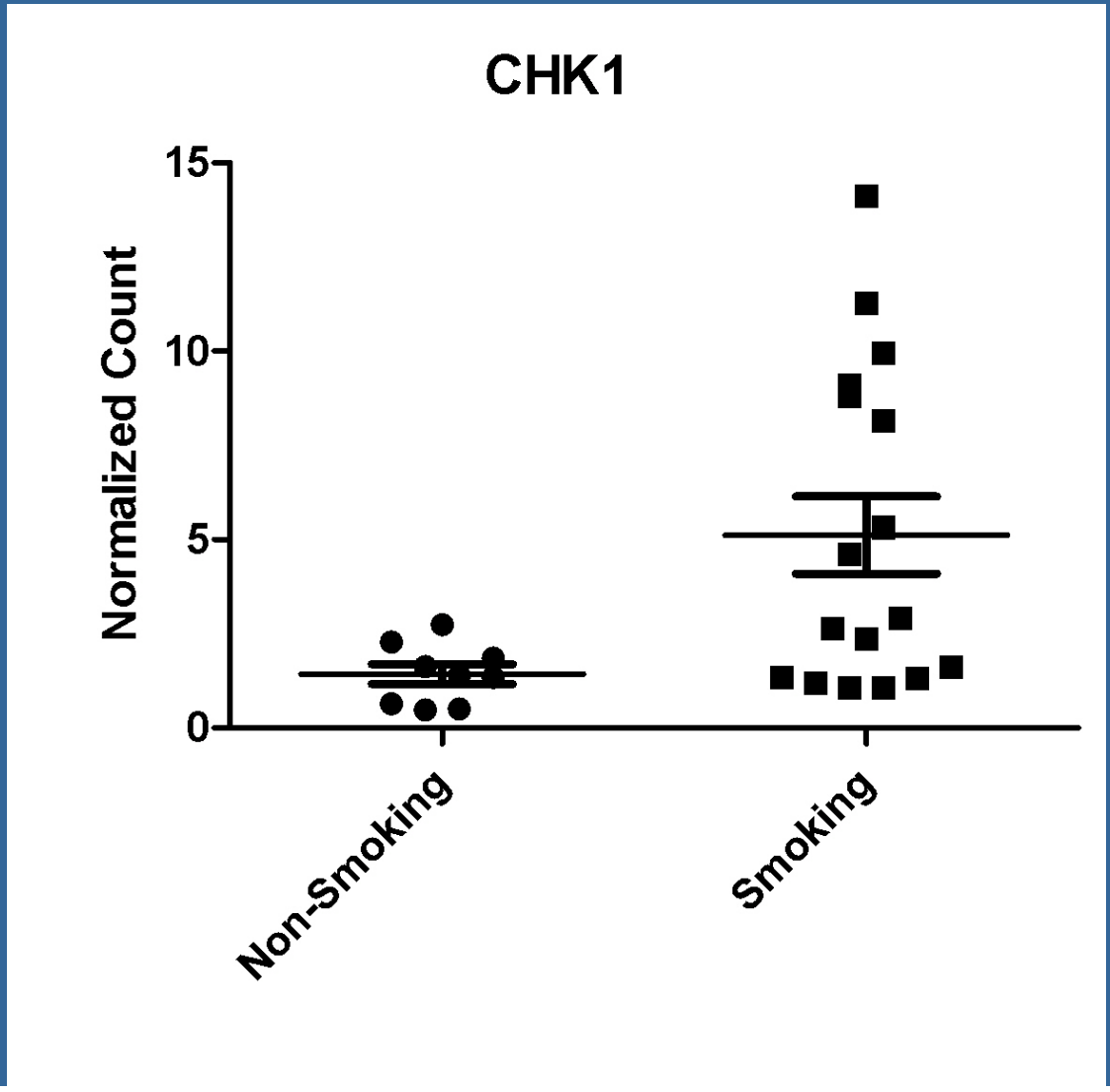




CHEK2


p=0.018
p=0.003
p=0.04
log2 ratio
5
4
3
2
1
0
HPV Positive Non-Smoker
HPV Positive Smoker
HPV Negative Non-Smoker
HPV Negative Smoker

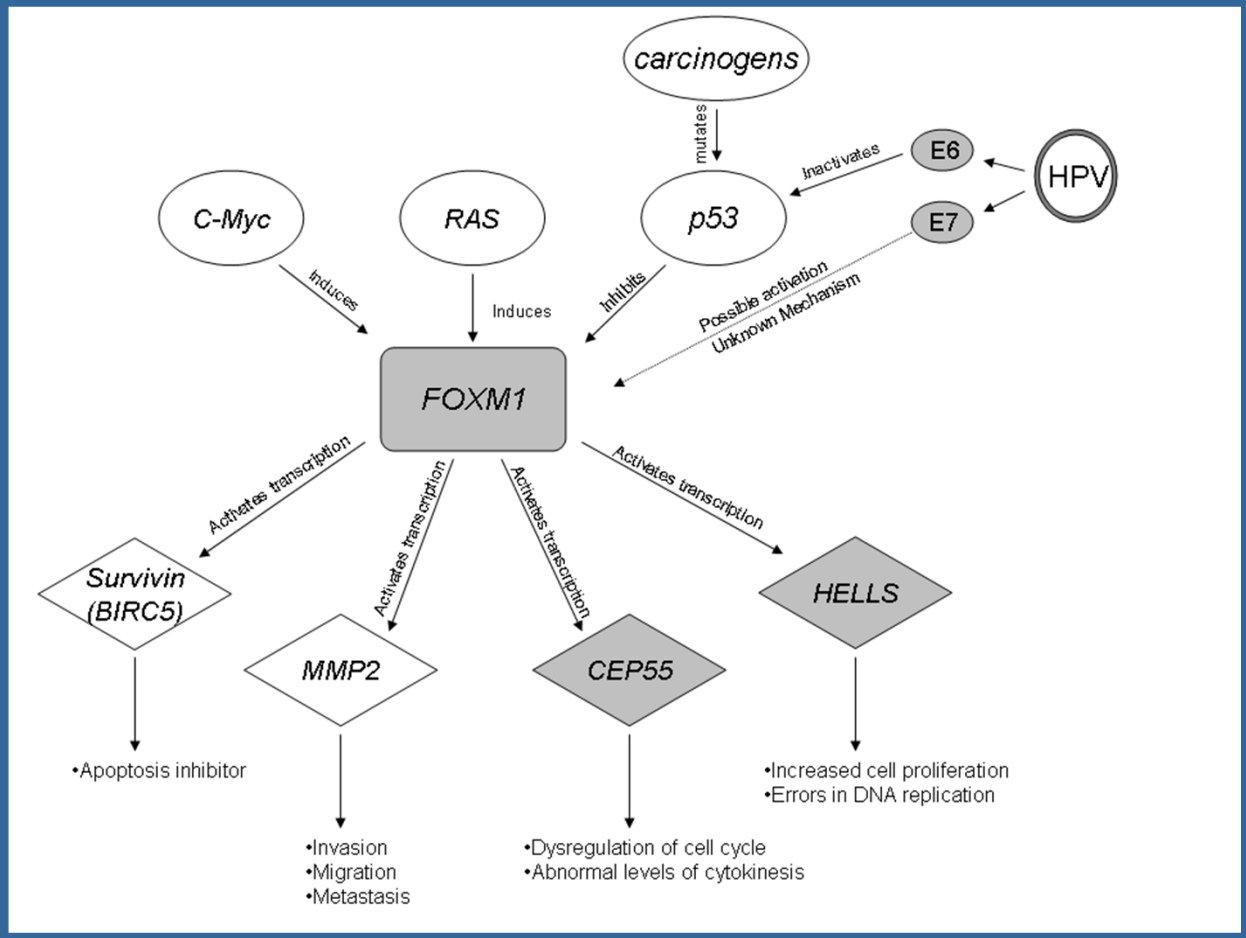
Variation in expression of Current Smoking groups

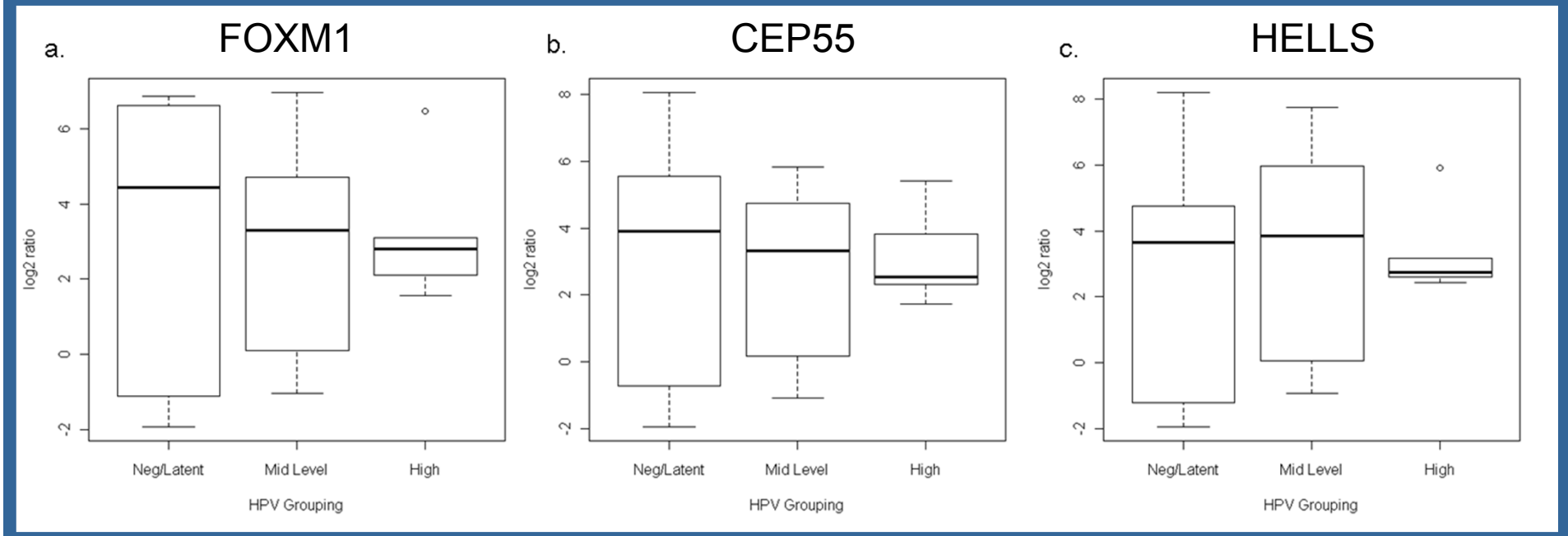


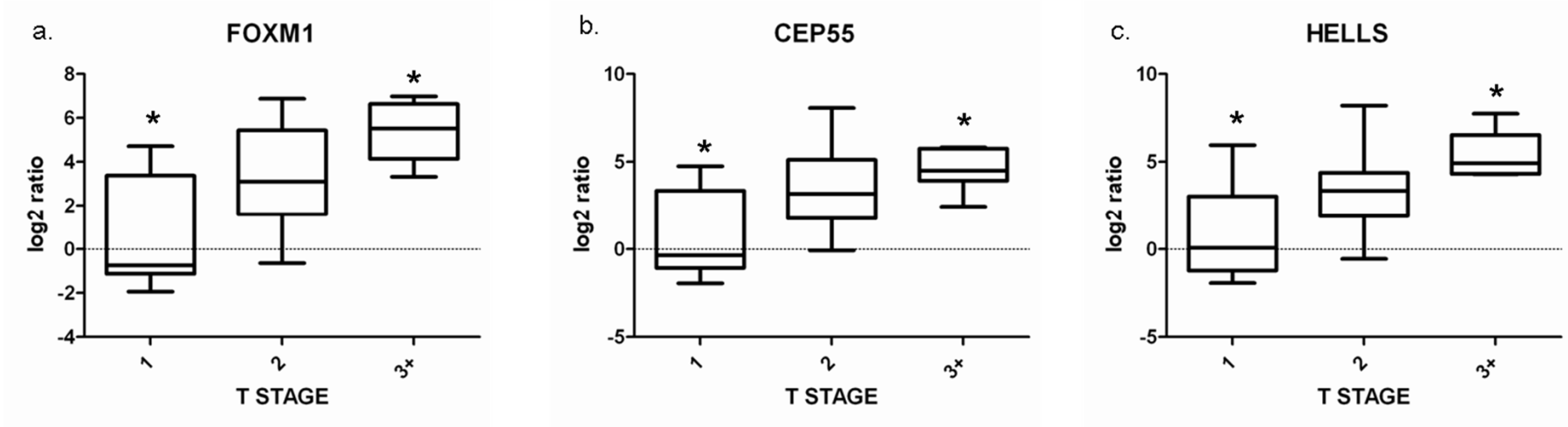
The Laryngoscope

Linking Expression of *FOXM1*, *CEP55* and *HELLS* to Tumorigenesis in Oropharyngeal Squamous Cell Carcinoma

Jeffrey R. Janus, MD; Rebecca R. Laborde, PhD; Alexandra J. Greenberg; Vivian W. Wang, PhD; Wei Wei, PhD; Anna Trier; Steven Olsen, MD; Eric J. Moore, MD; Kerry D. Olsen, MD; David I. Smith, PhD




a.
FOXM1
log2 ratio
Neg/Latent
Mid Level
High
HPV Grouping
b.
CEP55
log2 ratio
Neg/Latent
Mid Level
High
HPV Grouping
c.
HELLS
log2 ratio
Neg/Latent
Mid Level
High
HPV Grouping


a.
FOXM1
log2 ratio
T STAGE
b.
CEP55
log2 ratio
T STAGE
c.
HELLS
log2 ratio
T STAGE

What else is altered in cancer?

- Genome wide changes in methlylation
- Methylation changes may be a better direct indicator of environmental exposures
- How to measure methylation? Whole genome sequencing before and after bisulfite modification is one expensive way
- Alternative is a $450 array that measures 450,000 methylation sites in the genome. Downside- only a small fraction of all methylation sites

Methylation Array

- Infinium Human Methylation450 BeadChip Kits

	Age	Gender
Current Smoking	54	M
	56	M
	73	M
	65	M
	56	M
	41	M
	57	M
	50	M
Former Smoking	64	F
	48	M
	66	M
	69	M
	62	M
	56	M
	76	M
	59	M
Never Smoking	46	M
	73	M
	55	M
	54	F
	51	M
	58	M
	48	M
	37	M

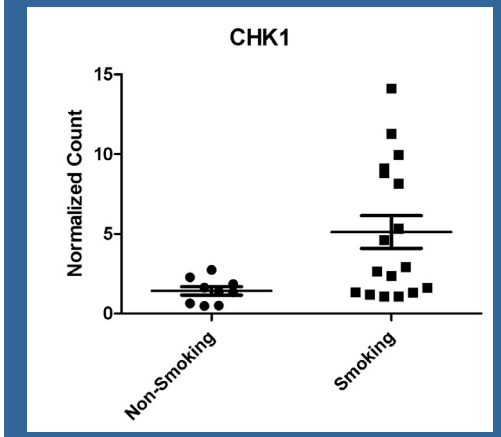
Background

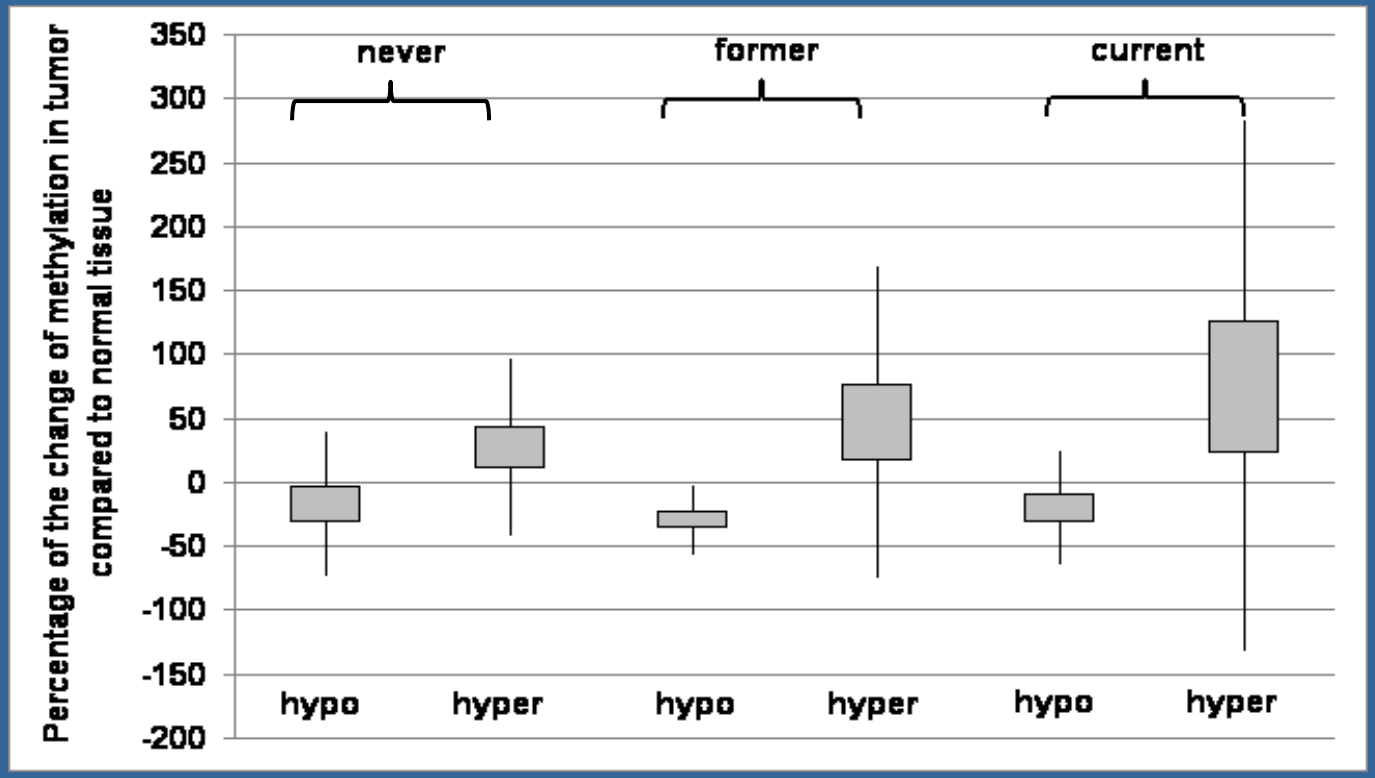
- Methylation: addition of methyl group to a substrate or the substitution of an atom or group by a methyl group.
- DNA methylation at CpG sites
  - conversion of the cytosine to 5-methylcytosine
  - Methylation in promoter region: Inverse relationship with transcriptional activity

Methylome

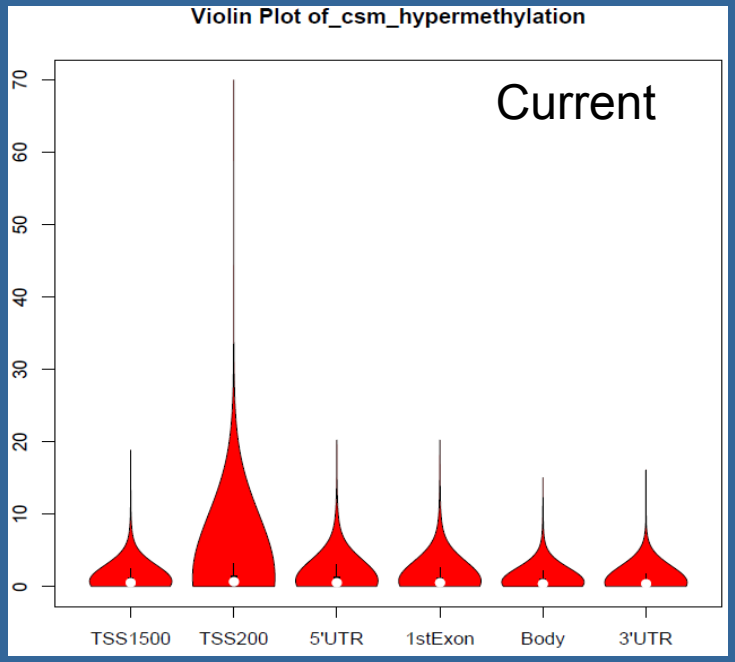
- Identify differentially methylated cg sites
  - SAM:
    - We chose a delta with which the median FDR among all cg sites is 0.01.
  - Limma:
    - Adjusted P-value < 0.01.
- Overlap of limma result and SAM result: 11,383 cg sites. 49,047 are hyper-methylated, 64,806 are hypo-methylated.

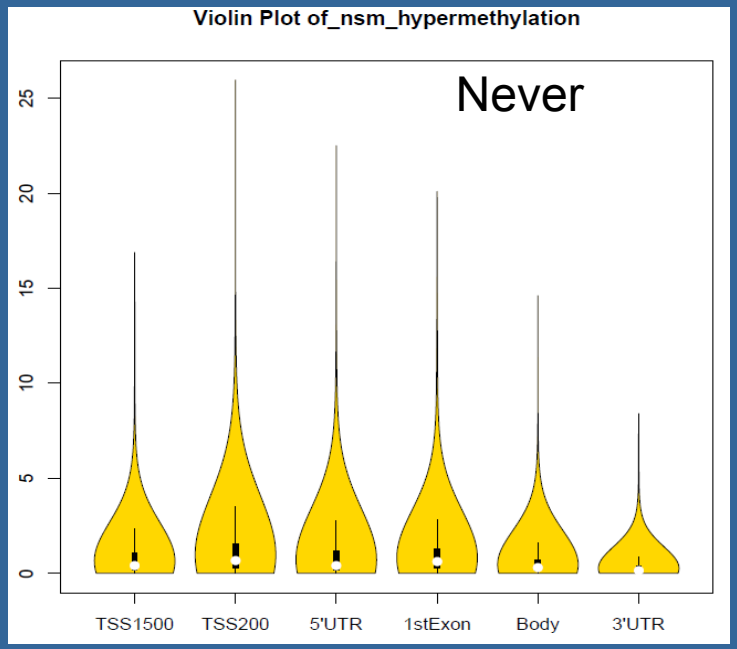
Similar patterns of variability when comparing transcriptome analysis and hypermethylation of current smoking patients

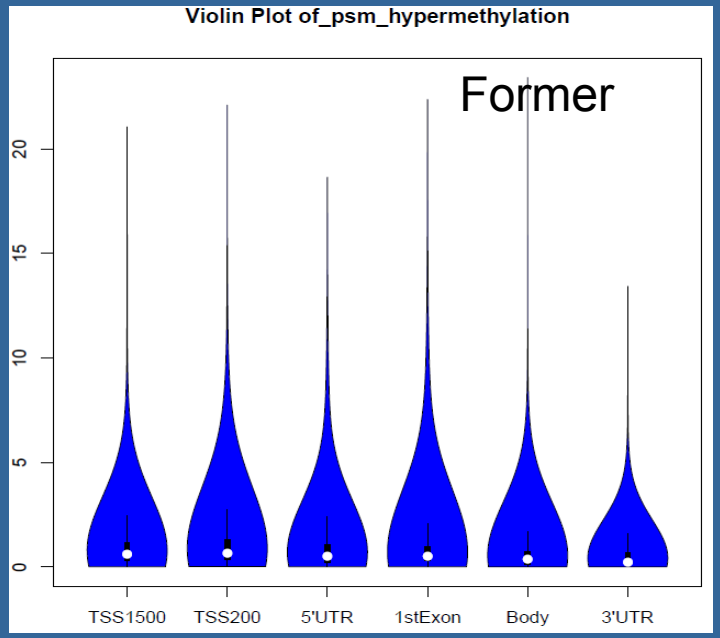




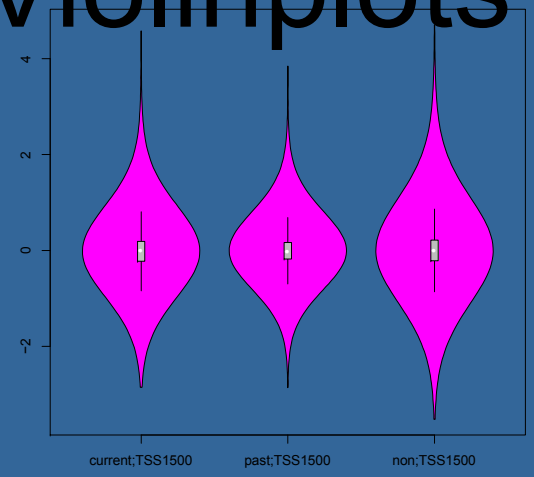
Hypermethylation of specific gene segments

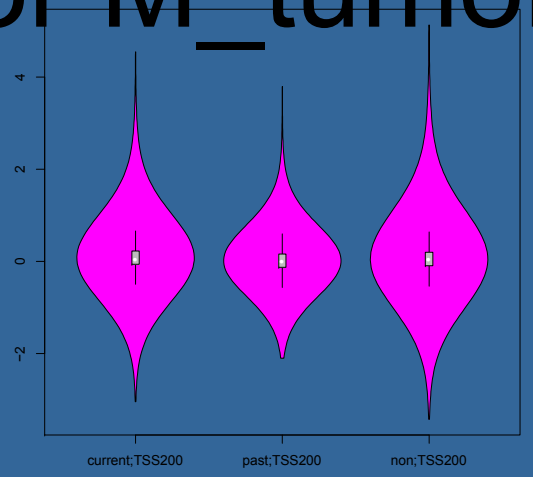


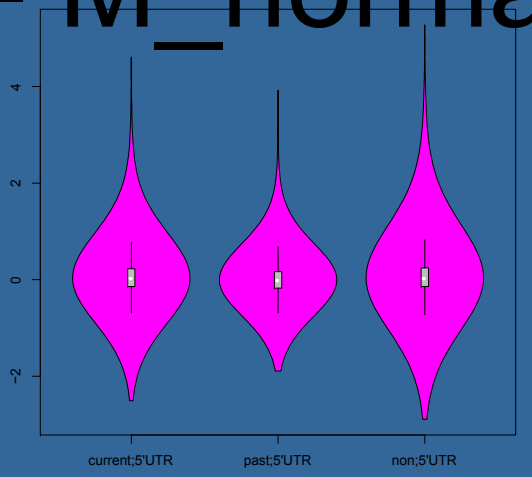




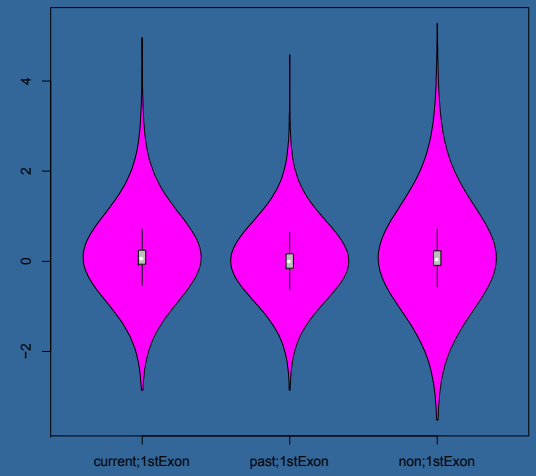
Violinplots for M_tumor - M_normal

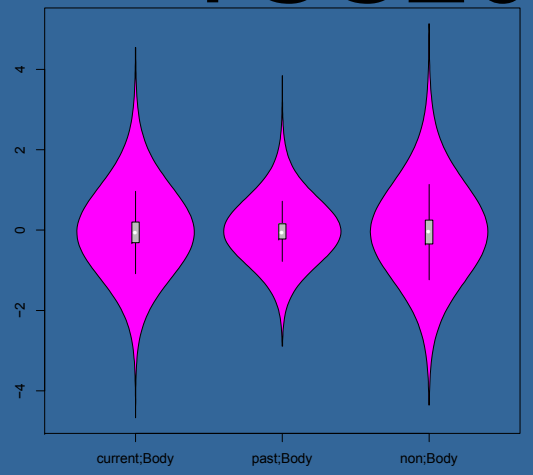


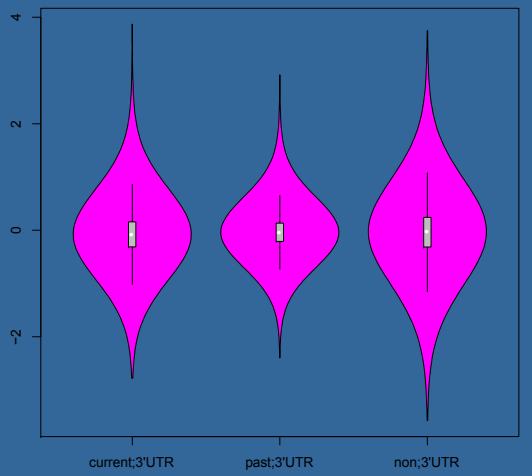




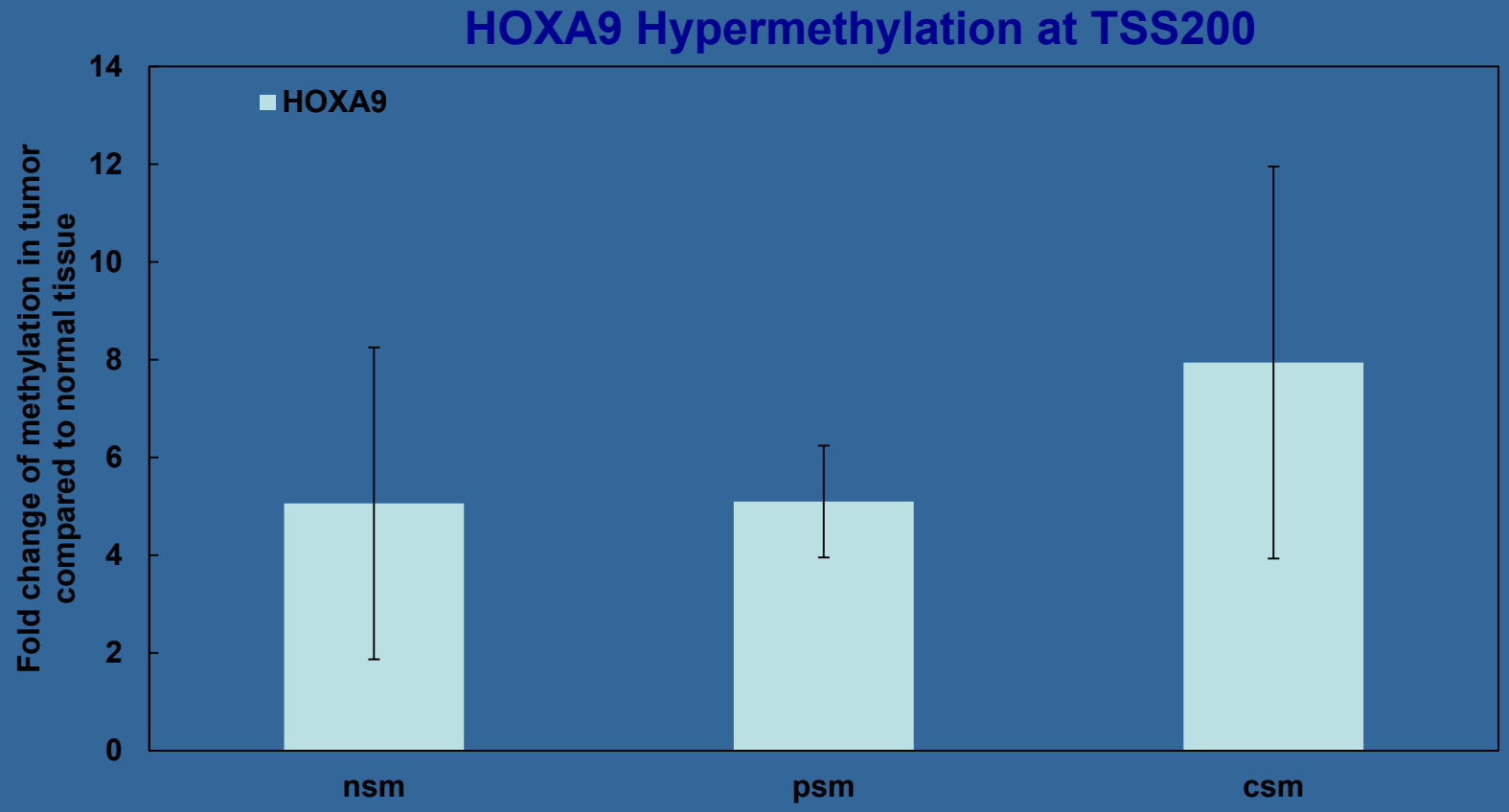
TSS200

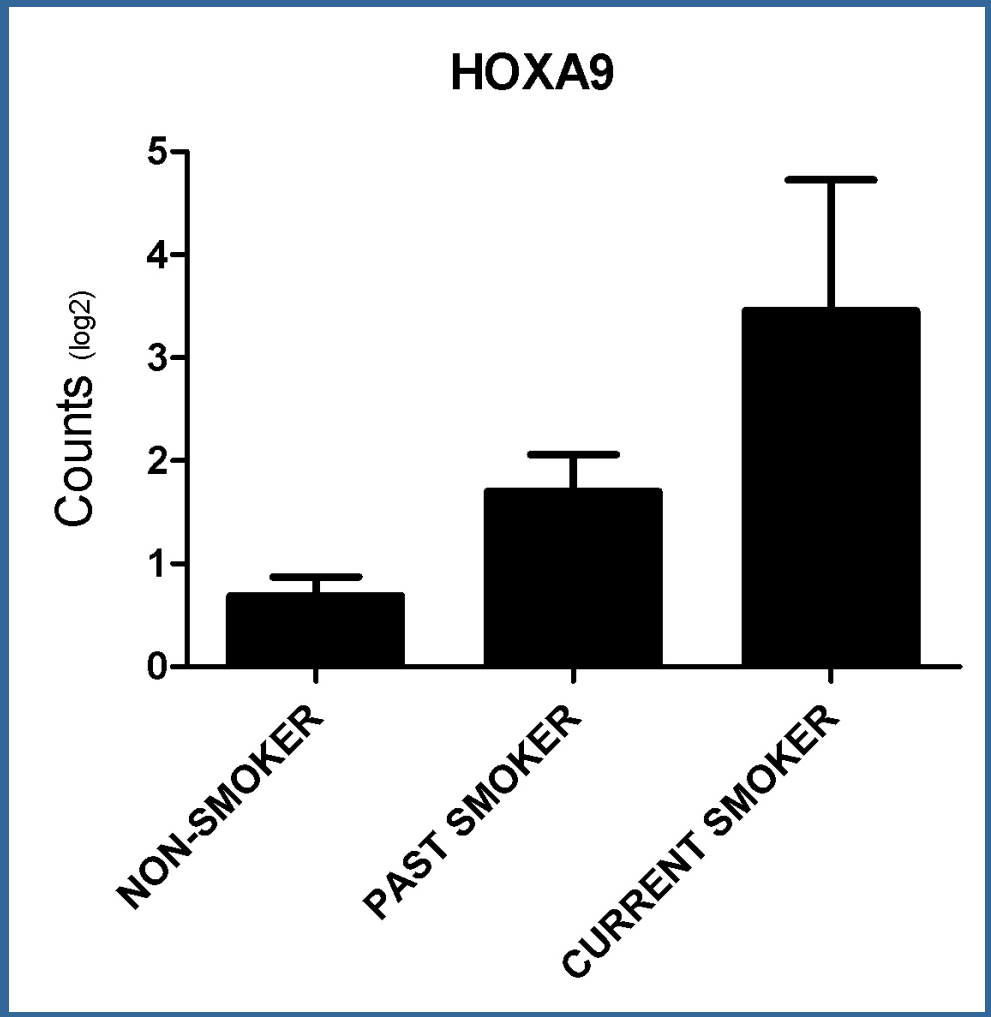






Body


HOXA9 Hypermethylation at TSS200
HOXA9
Fold change of methylation in tumor compared to normal tissue
14
12
10
8
6
4
2
0
nsm
psm
csm


HOXA9
Counts (log2)
5
4
3
2
1
0
NON-SMOKER
PAST SMOKER
CURRENT SMOKER

Correlation Between Two Datasets

- 7 patients have both methylome and transcriptome data.

- Spearman correlation:
  - 14 pairs of M-value and expression value
  - cg site and the gene it lies on
  - Hyper-methylation with down-regulation; hypo-methylation with up-regulation.

Two Lists of Genes Chosen

- Genes that are differentially methylated when comparing tumor to normal and have corresponding expression change
- Genes that are differentially methylated when comparing the smokers to the non-smokers and ex-smokers

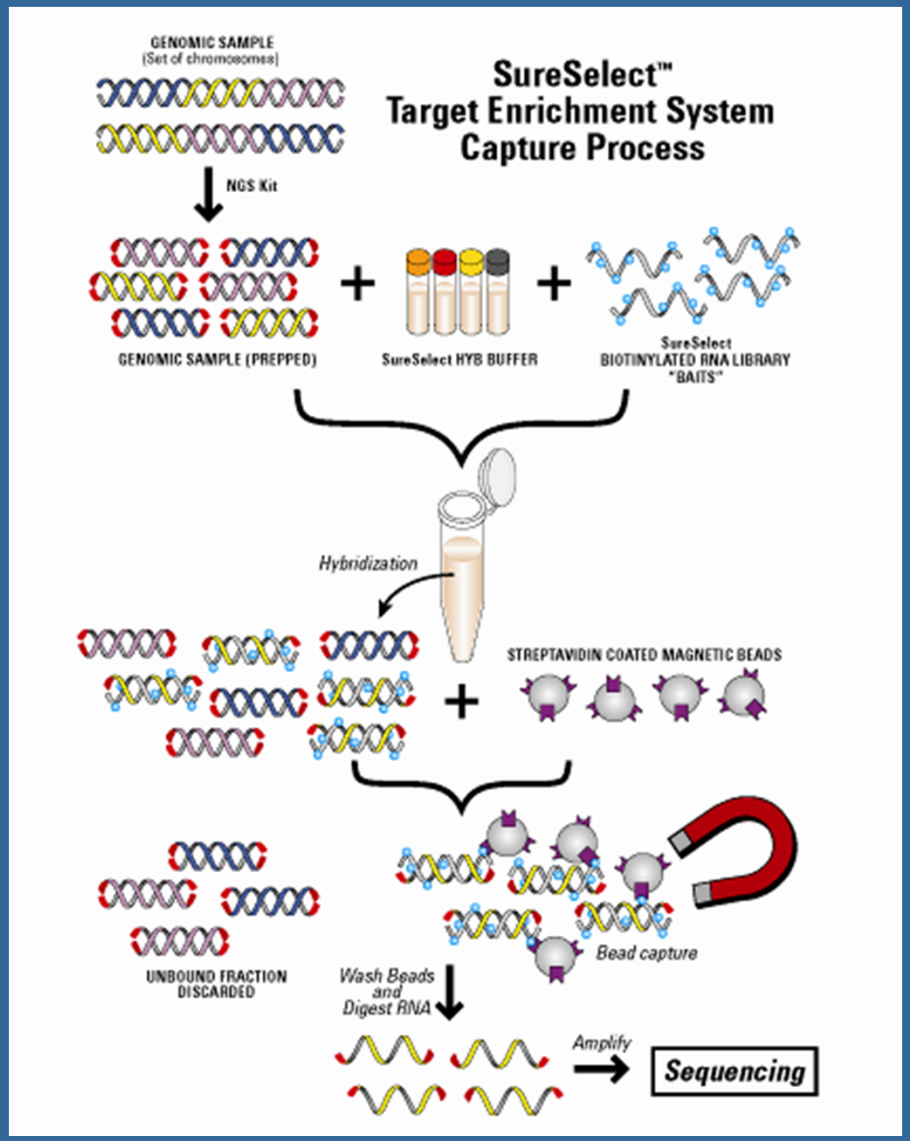
Validation

- Limited resources (both financial and in terms of tissue material)
- Fluidigm- Nanofluidics solution that can measure expression of 96 genes (or 48 in duplicate) against 96 samples
- Requires 5 nl per individual reaction
- Both cost and tissue effective- but you still need to purchase TaqMan probes for each gene

Whole Exome Sequencing

- Instead of sequencing the entire 6 Gb genome, why not “pull down” the 38 Mb exome?
- Much less sequencing and can sequence the exome at depth (100+ coverage)

SureSelect™ Target Enrichment System: Workflow



Agilent Technologies
SureSelect™
Target Enrichment System

Exome Sequencing

- Illumina TruSeq Library and Exome Enrichment kit

Sample	Age	Smoking	HPV
1	54	Current	Neg
2	46	Current	Neg
3	56	Current	Neg
4	69	Current	Neg
5	37	Never	Positive
6	57	Never	Positive
7	48	Never	Positive
8	54	Never	Positive

Variant Summary

		Smokers				Non-smokers			
		463	527	691	718	663	687	716	721
	Total SNVs	117,902	144,643	142,491	133,823	141,016	114,960	144,270	141,163
	Common SNVs	114,195	142,578	140,236	124,940	136,549	106,287	141,058	137,453
Tumor-specific variants	Total	3,707	2,065	2,255	8,883	4,467	8,673	3,212	3,710
	non-synonymous	179	69	72	345	171	414	122	191
	non (novel)	81	25	25	71	48	106	37	65
	% novel	45.3%	36.2%	34.7%	20.6%	28.1%	25.6%	30.3%	34.0%
Tumor-specific mutations	Total	1,000	519	3,350	824	930	590	409	662
	non-synonymous	109	34	34	59	74	34	43	85
	non (novel)	69	16	18	40	36	15	25	56
	% novel	63.3%	47.1%	52.9%	67.8%	48.6%	44.1%	58.1%	65.9%

Tumor-specific variant = tumor sample different than normal (most common = LOH in tumor)
Tumor-specific mutation = same as reference in normal sample, variant in cancer sample

Novel = not in dbsnp132

Patient 463 (Current smoker, Low HPV)

Total: 117,902 snvs
Common to T and N: 114,195
Tumor-specific variants: 3707

Non-synonymous: 179

Pathway	Genes	KEGG	Totals List	Totals Gene Set	z-score
Metabolic pathways			5	967	-2.72
Leishmaniasis			3	64	2.13
Viral myocarditis			3	65	2.10
Allograft rejection			2	30	2.36
Asthma			2	24	2.78
Base excision repair			2	27	2.55
Graft-versus-host disease			2	31	2.30
Intestinal immune network for IgA production			2	35	2.08
Type I diabetes mellitus			2	36	2.03

Enriched pathways

KEGG Export

Chrom.	Position	Gene	Region	Type	dbSNP	Exon Pop	Low Cov Pop	Ref.	6747: Var +	6748: Var +
chr1	2,444,414	PANK4	CDS	snv	rs7535528	0	0	G	R	A
chr1	15,438,990	RP1-21O18.1, C1orf126	CDS	snv	rs10803354	0	0	G	A	R
chr1	45,797,505	MUTYH	CDS	snv	rs3219489	0	0	C	G	S
chr1	144,879,264	PDE4DIP	CDS	snv	rs2798901	0	0	A	R	G
chr1	160,851,826	ITLN1	CDS	snv	rs2274907	0	0	A	W	T
chr1	248,085,011	OR2T8	CDS	snv	rs4595394	0	0	C	S	G
chr1	248,801,778	OR2T35	CDS	snv	rs61834268	0	0	A	W	T
chr2	42,991,200	OXER1	CDS	snv	rs34142793	0	0	C	S	G
chr3	49,156,473	USP19	CDS	snv	rs11552724	0	0	C	S	G
chr3	49,940,078	MST1R	CDS	snv	rs2230593	0	0	C	Y	T
chr3	75,719,154	LOC100506764	CDS	snv	-	0	0	C	M	A
chr3	195,510,073	MUC4	CDS	snv	-	0	0	G	R	A
chr4	7,436,239	SORCS2, PSAPL1	CDS	snv	rs56402179	0	0	C	Y	T
chr4	8,583,312	GPR78	CDS	snv	rs17844778	0	0	A	M	C
chr4	40,428,010	RBM47	CDS	snv	rs278981	0	0	T	C	Y
chr5	70,806,958	BDP1	CDS	snv	rs6886336	0	0	G	A	R
chr6	31,838,441	SLC44A4	CDS	snv	rs114483117, rs644827	0	0	T	Y	C
chr6	32,020,717	TNXB	CDS	snv	-	0	0	G	K	T
chr6	197,943	HLA-DRB5	CDS	snv	rs17841043, rs74243539	0	0			G

Patient 463 (Current smoker, Low HPV)

Total: 117,902 snvs
Common to T and N: 114,195
Tumor-specific mutations: 1000
Non-synonymous: 109
Not in dbSNP: 69

Pathway	Genes	KEGG	Totals List	Totals Gene Set	z-score
Viral myocarditis			3	65	2.47
ABC transporters			2	43	2.03
Allograft rejection			2	30	2.69
Asthma			2	24	3.14
Autoimmune thyroid disease			2	37	2.30
Graft-versus-host disease			2	31	2.63
Intestinal immune network for IgA production			2	35	2.40
Type I diabetes mellitus			2	36	2.35

Enriched pathways

Tumor **Normal**

Chrom.	Position	Gene	Region	Type	dbSNP	Exon Pop	Low Cov Pop	Ref.	6747: Var +	6748: Var +
chr1	6,278,414	RNF207	CDS	snv	rs709209	0	0	A	R	NC
chr1	12,854,414	PRAMEF1	CDS	snv	rs1063769	0	0	G	R	NC
chr1	12,907,781	HNRNPCL1	CDS	snv	rs1737105, rs28441396	0	0	T	Y	NC
chr1	12,907,798	HNRNPCL1	CDS	snv	-	0	0	A	M	NC
chr1	12,907,802	HNRNPCL1	CDS	snv	-	0	0	C	Y	NC
chr1	12,907,803	HNRNPCL1	CDS	snv	-	0	0	C	Y	NC
chr1	16,902,894	NBPF1	CDS	snv	rs61772339	0	0	A	R	NC
chr1	24,407,877	MYOM3	CDS	snv	-	0	0	C	Y	NC
chr1	40,960,924	ZNF642	CDS	snv	-	0	0	G	S	NC
chr1	43,149,102	YBX1	CDS	snv	-	0	0	G	R	NC
chr1	45,484,178	ZSWIM5	CDS	snv	-	0	0	A	W	NC
chr1	144,854,581	PDE4DIP	CDS	snv	rs78371650	0	0	T	Y	NC
chr1	152,276,626	FLG	CDS	snv	rs3126075	0	0	G	S	NC
chr1	181,727,945	CACNA1E	CDS	snv	-	0	0	G	R	NC
chr1	230,513,309	PGBD5	CDS	snv	-	0	0	C	Y	NC
chr1	248,524,992	OR2T4	CDS	snv	rs45552134	0	0	A	R	NC
chr1	248,651,959	OR2T5	CDS	snv	rs76332972	0	0	C	Y	NC
chr2	10,263,618	RRM2	CDS	snv	-	0	0	G	K	NC
chr2	[illegible]	C2orf70	[illegible]	snv	-	0	0	A	R	NC

Patient 687 (Non Smoker, High HPV)

Total: 114,960 snvs
Common to T and N: 106,287
Tumor-specific variants: 8,673
Non-synonymous: 414
Not in dbSNP: 106

Pathway	Genes	KEGG	Totals List	Totals Gene Set	z-score
Allograft rejection			4	31	3.98
Graft-versus-host disease			4	33	3.80
Type I diabetes mellitus			4	38	3.42
Autoimmune thyroid disease			4	40	3.29
p53 signaling pathway			5	62	3.07
Phosphatase and tensin homolog					
Thrombospondin 1					
Tumor protein p53					
Tumor protein p73					
PERP, TP53 apoptosis effector					
Viral myocarditis			5	64	2.99
Endometrial cancer			4	51	2.68
Inositol phosphate metabolism			4	56	2.45
Antigen processing and presentation			4	57	2.41
Fatty acid biosynthesis			1	6	2.36
Phagosome			7	133	2.34
Bladder cancer			3	39	2.28
RNA transport			7	140	2.19
Asthma			2	22	2.15
Natural killer cell mediated cytotoxicity			6	118	2.07
ABC transporters			3	44	2.03

Enriched pathways

KEGG Export

Chrom.	Position	Gene	Region	Type	dbSNP	Exon Pop	Low Cov Pop	Ref.	T: Var +	N: Var +
chr1	1,849,529	TMEM52	CDS	snv	rs28640257	0	0	A	G	R
chr1	2,458,010	PANK4	CDS	snv	-	0	0	G	S	C
chr1	2,488,153	LOC100133445, TNFRSF14	CDS	snv	rs4870	0	0	A	G	R
chr1	6,279,370	RNF207	CDS	snv	rs846111	0	0	G	C	S
chr1	6,614,535	NOL9	CDS	snv	rs4908923	0	0	G	A	R
chr1	9,323,910	H6PD	CDS	snv	rs6688832	0	0	G	A	R
chr1	11,736,959	MAD2L2	CDS	snv	-	0	0	C	T	Y
chr1	12,009,956	PLOD1	CDS	snv	rs7551175	0	0	G	A	R
chr1	12,921,110	PRAMEF2	CDS	snv	rs12139546	0	0	A	G	R
chr1	12,921,132	PRAMEF2	CDS	snv	rs12139550	0	0	A	G	R
chr1	12,942,119	PRAMEF4	CDS	snv	rs75355616	0	0	C	Y	T
chr1	12,942,961	PRAMEF4	CDS	snv	rs4625290	0	0	A	C	M
chr1	13,036,587	PRAMEF22	CDS	snv	rs76348406, rs77788641	0	0	C	T	Y
chr1	15,808,767	CELA2B	CDS	snv	rs3820071	0	0	G	A	R
chr1	15,808,872	CELA2B	CDS	snv	rs3766160	0	0	G	A	R
chr1	16,332,665	C1orf64	CDS	snv	rs34950166	0	0	C	T	Y
chr1	16,532,498	ARHGEF19	CDS	snv	rs41269185	0	0	G	A	R
chr1	16,577,908	FBXO42	CDS	snv	rs12069239	0	0	G	C	S

Patient 687 (Non Smoker, High HPV)

Total: 114,960 snvs
Common to T and N: 106,287
Tumor-specific mutations: 943
Non-synonymous: 106

Pathway	Genes	KEGG	Totals List	Totals Gene Set	z-score
Hematopoietic cell lineage			3	76	2.12
Valine, leucine and isoleucine biosynthesis			1	10	2.48

Enriched pathways

KEGG Export

Chrom.	Position	Gene	Region	Type	dbSNP	Exon Pop	Low Cov Pop	Ref.	T: Var +	N: Var +
chr1	3,648,111	TP73	CDS	snv	-	0	0	G	R	NC
chr1	9,099,992	SLC2A5	CDS	snv	-	0	0	C	Y	NC
chr1	12,921,485	PRAMEF2	CDS	snv	-	0	0	A	R	NC
chr1	16,260,450	SPEN	CDS	snv	-	0	0	C	Y	NC
chr1	16,862,472	LOC100506533	CDS	snv	rs2779468	0	0	G	S	NC
chr1	32,694,187	EIF3I	CDS	snv	-	0	0	G	R	NC
chr1	89,449,483	CCBL2, RBMXL1	CDS	snv	rs74100106	0	0	C	S	NC
chr1	144,619,903	NBPF9	CDS	snv	rs71527379, rs7525482	0	0	T	W	NC
chr1	144,621,643	NBPF9	CDS	snv	rs11488527	0	0	G	R	NC
chr1	145,508,934	RBM8A	CDS	snv	-	0	0	T	K	NC
chr1	152,882,979	IVL	CDS	snv	rs17855670	0	0	C	Y	NC
chr1	207,749,010	CR1	CDS	snv	-	0	0	C	Y	NC
chr1	214,814,146	CENPF	CDS	snv	-	0	0	C	Y	NC
chr2	18,766,138	NT5C1B	CDS	snv	-	0	0	G	R	NC
chr2	39,222,390	SOS1	CDS	snv	-	0	0	C	Y	NC
chr2	55,646,019	CCDC88A	CDS	snv	-	0	0	C	Y	NC
chr2	96,517,503	LOC400986	CDS	snv	rs79524470	0	0	A	R	NC
chr2	98,129,707	ANKRD36B	CDS	snv	rs5011467	0	0	G	K	NC
		36B	CDS	snv	rs24639	0	0		R	NC

Genes mutated in at least 3 out 4 patients

Non-smokers

IGSF3*

CDC27

MUC4

Smokers

CDC27

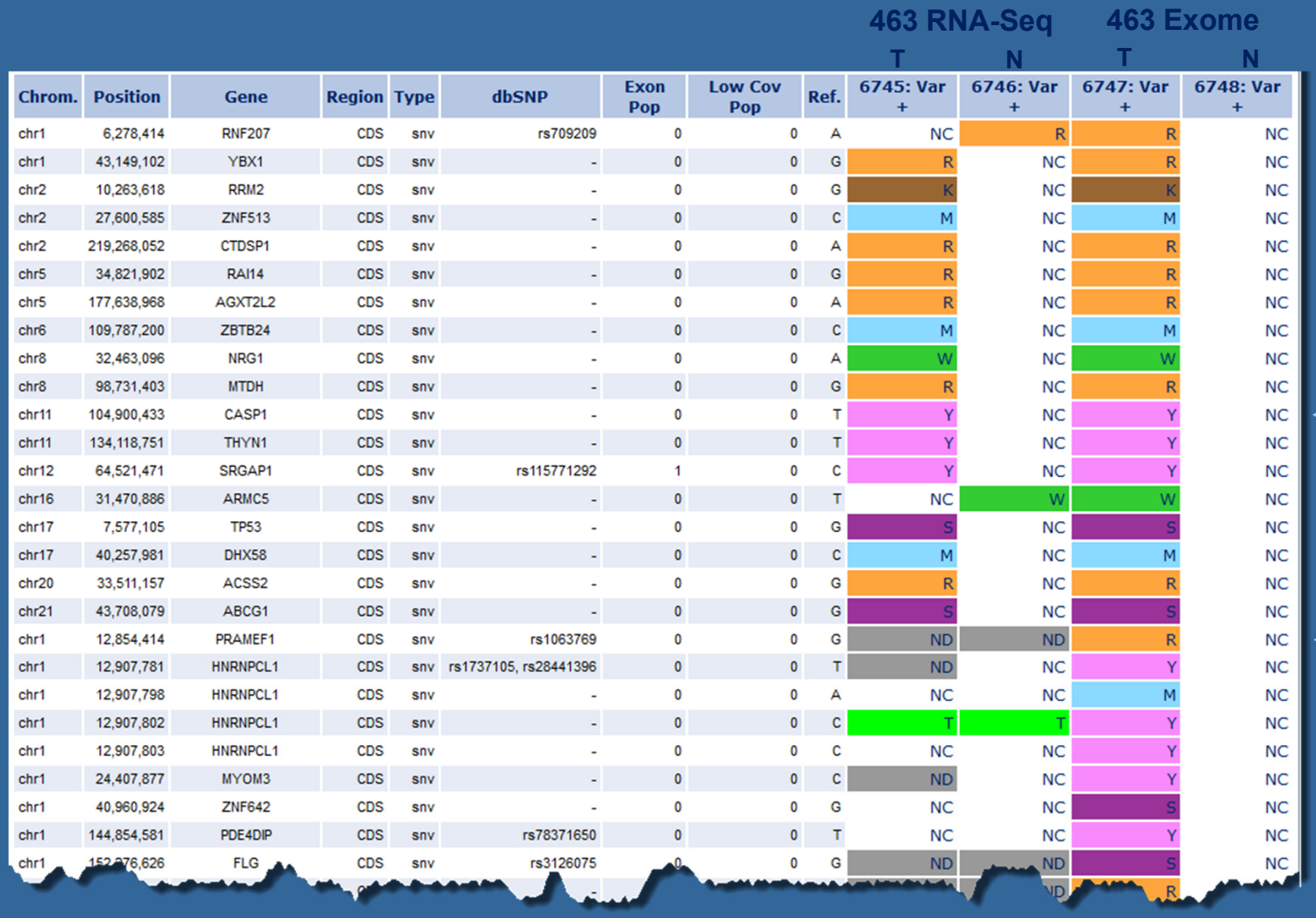
LOC100289375

MUC4*

MUC6

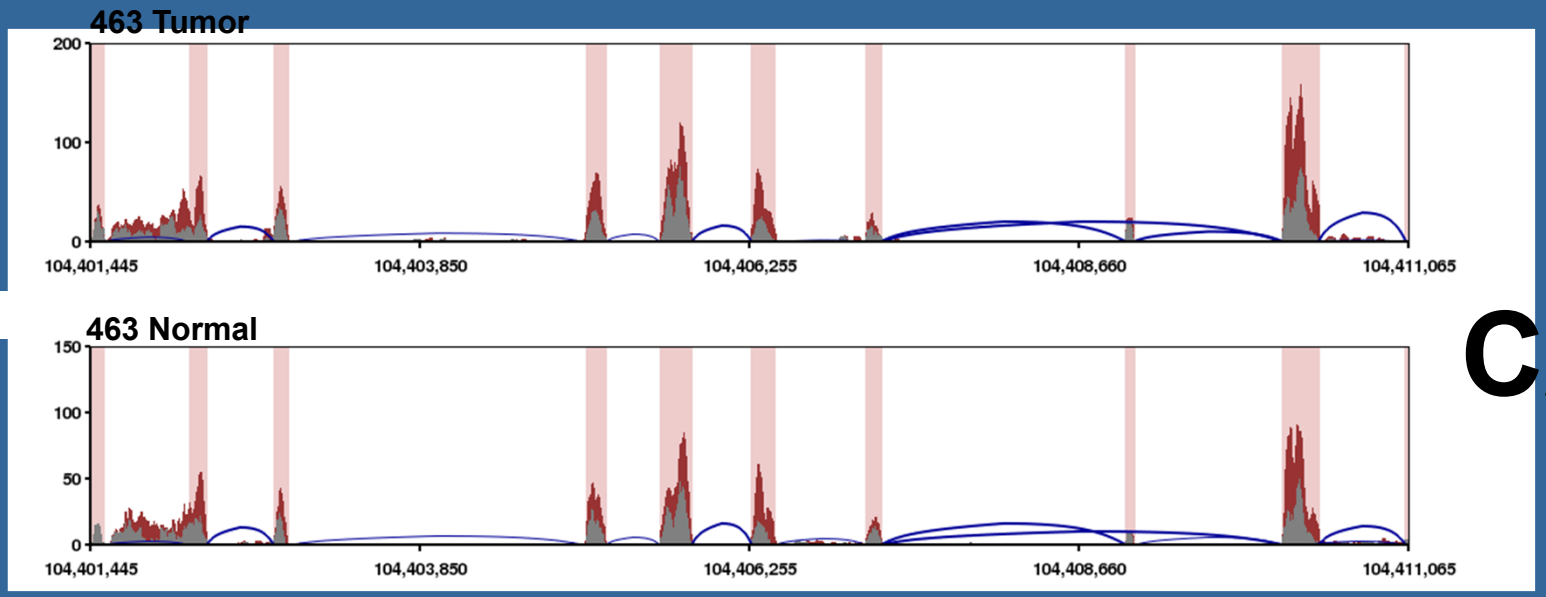
PRSS3

Combining Exome-seq and RNA-Seq

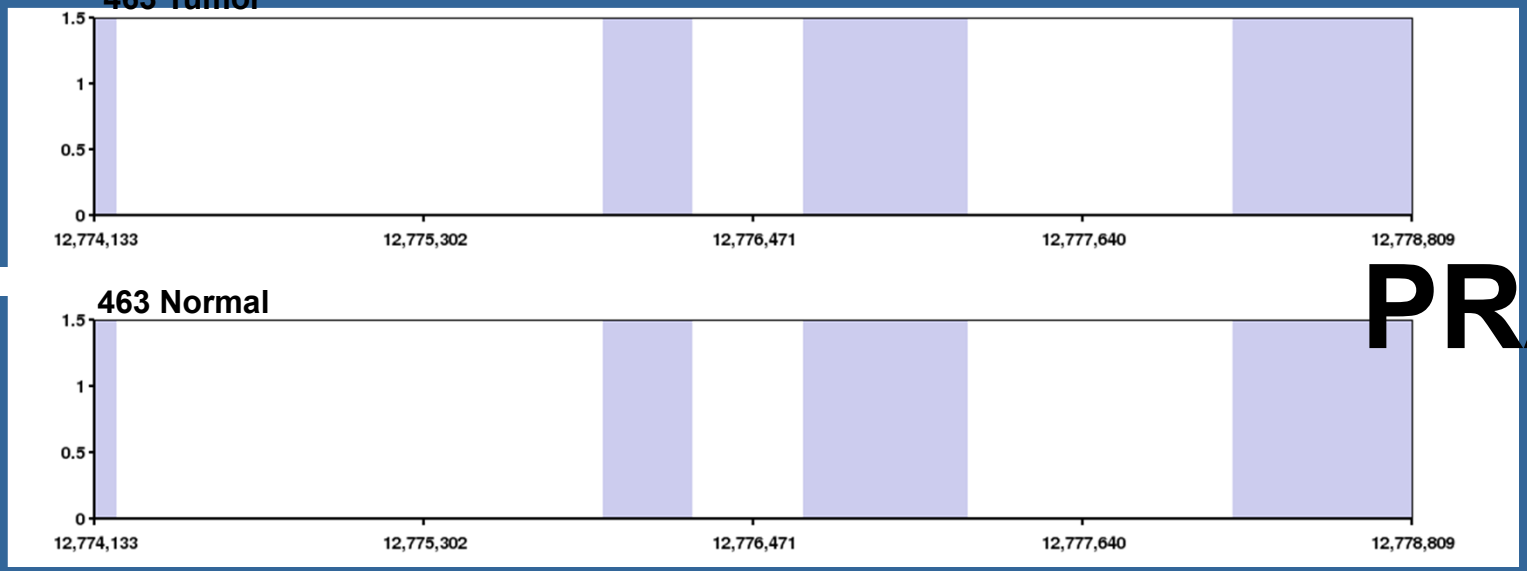


									463 RNA-Seq		463 Exome		
									T	N	T	N	
Chrom.	Position	Gene	Region	Type	dbSNP	Exon Pop	Low Cov Pop	Ref.	6745: Var +	6746: Var +	6747: Var +	6748: Var +	
chr1	6,278,414	RNF207	CDS	snv	rs709209	0	0	A	NC	R	R	NC	
chr1	43,149,102	YBX1	CDS	snv	-	0	0	G	R	NC	R	NC	
chr2	10,263,618	RRM2	CDS	snv	-	0	0	G	K	NC	K	NC	
chr2	27,600,585	ZNF513	CDS	snv	-	0	0	C	M	NC	M	NC	
chr2	219,268,052	CTDSP1	CDS	snv	-	0	0	A	R	NC	R	NC	
chr5	34,821,902	RAI14	CDS	snv	-	0	0	G	R	NC	R	NC	
chr5	177,638,968	AGXT2L2	CDS	snv	-	0	0	A	R	NC	R	NC	
chr6	109,787,200	ZBTB24	CDS	snv	-	0	0	C	M	NC	M	NC	
chr8	32,463,096	NRG1	CDS	snv	-	0	0	A	W	NC	W	NC	
chr8	98,731,403	MTDH	CDS	snv	-	0	0	G	R	NC	R	NC	
chr11	104,900,433	CASP1	CDS	snv	-	0	0	T	Y	NC	Y	NC	← CASP1
chr11	134,118,751	THYN1	CDS	snv	-	0	0	T	Y	NC	Y	NC	
chr12	64,521,471	SRGAP1	CDS	snv	rs115771292	1	0	C	Y	NC	Y	NC	
chr16	31,470,886	ARMC5	CDS	snv	-	0	0	T	NC	W	W	NC	
chr17	7,577,105	TP53	CDS	snv	-	0	0	G	S	NC	S	NC	
chr17	40,257,981	DHX58	CDS	snv	-	0	0	C	M	NC	M	NC	
chr20	33,511,157	ACSS2	CDS	snv	-	0	0	G	R	NC	R	NC	
chr21	43,708,079	ABCG1	CDS	snv	-	0	0	G	S	NC	S	NC	
chr1	12,854,414	PRAMEF1	CDS	snv	rs1063769	0	0	G	ND	ND	R	NC	← PRAMEF1
chr1	12,907,781	HNRNPCL1	CDS	snv	rs1737105, rs28441396	0	0	T	ND	NC	Y	NC	
chr1	12,907,798	HNRNPCL1	CDS	snv	-	0	0	A	NC	NC	M	NC	
chr1	12,907,802	HNRNPCL1	CDS	snv	-	0	0	C	T	T	Y	NC	
chr1	12,907,803	HNRNPCL1	CDS	snv	-	0	0	C	NC	NC	Y	NC	
chr1	24,407,877	MYOM3	CDS	snv	-	0	0	C	ND	NC	Y	NC	
chr1	40,960,924	ZNF642	CDS	snv	-	0	0	G	NC	NC	S	NC	
chr1	144,854,581	PDE4DIP	CDS	snv	rs78371650	0	0	T	NC	NC	Y	NC	
chr1	152,276,626	FLG	CDS	snv	rs3126075	0	0	G	ND	ND	S	NC	
					-					ND	R		

ND = No Data

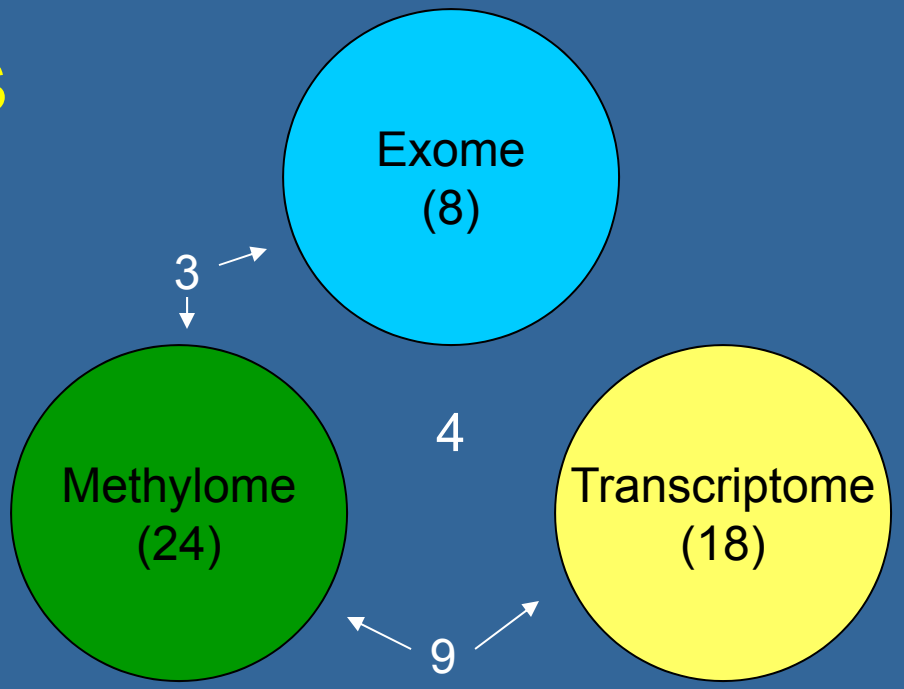

463 Tumor
200
100
0
104,401,445
104,403,850
104,406,255
104,408,660
104,411,065
463 Normal
150
100
50
0
104,401,445
104,403,850
104,406,255
104,408,660
104,411,065

CASP


463 Tumor
1.5
1
0.5
0
12,774,133
12,775,302
12,776,471
12,777,640
12,778,809
463 Normal
1.5
1
0.5
0
12,774,133
12,775,302
12,776,471
12,777,640
12,778,809

PRAME

Integrating datasets



- Correlate patterns of differential expression to
  - DNA Mutations
  - Differential Methylation
- Validate exome mutations in transcriptome
- Identify RNA editing events
- Correlate methylation patterns with differential splicing events

Moving forward

- Identify gene expression patterns of key pathways that associate with risk factor groups
  - In an individual patient, which risk factors are most significantly influencing disease development
- Compare multiple datasets to determine mechanism
  - Exome sequencing (mutations)
  - Methylation Arrays
  - Biological marker of tobacco exposure

General Conclusions

- Sequencing tools have important potential for clinical application
- Importance of building larger datasets
  - Correlate with disease outcome as they mature
- Importance of integrating datasets
    - Correlate with exposure, etiology, response to therapy
- Bottleneck involves bioinformatic challenges

Acknowledgements

Rebecca Laborde, Ph.D
Vivian Wang, Ph.D
Terra Lasho

Analysis
Yan Asmann, Ph.D
Sumit Middha
Wei Wei, Ph.D (UMR)

Geospiza
Todd Smith, Ph.D
Eric Olsen, PhD

UIUC
Jian Ma, Ph.D.
Yiyi Liu

Otolaryngology
Charles Beatty, MD

Kerry Olson, MD
Jan Kasperbauer, MD
Eric Moore, MD

Steven Olson, MD
Jeff Janus, MD

Nicole Tombers

Anatomic Pathology
Joaquin Garcia, MD