Mutation Analysis in Frozen and FFPE Tumor Samples

Gad Getz, PhD
Kristin Ardlie, PhD

Broad Institute of Harvard and MIT

BROAD INSTITUTE

Why use FFPE?

- Very large numbers of samples in tissue banks and Biorepositories worldwide

- Samples often very well-characterized with histological, pathological and follow-up clinical data

➔ Can fill the accrual gap in TCGA (and future of TCGA)
"We need to get to 10,000 patients per tumor type" -- Lou Staudt (Nov 2012)

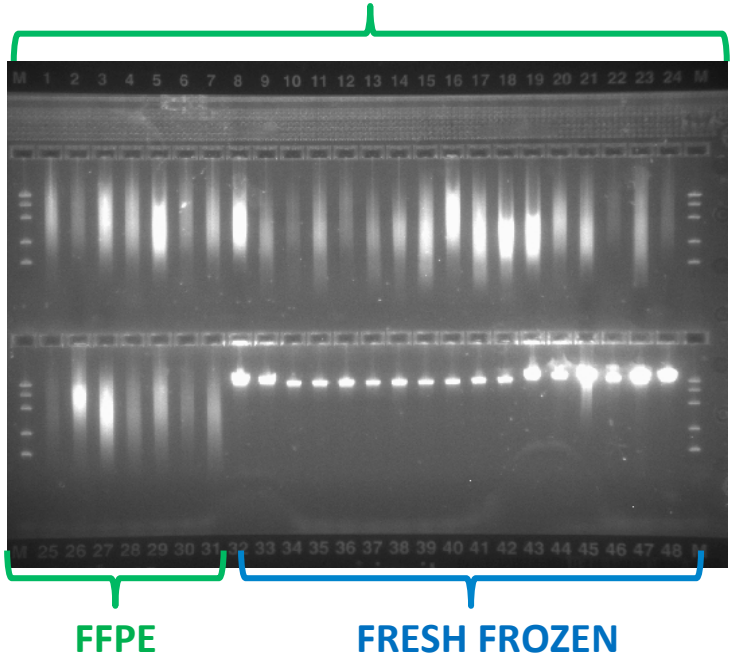
- Remains part of clinical standard of care (difficult to change pathology practices for research needs alone)

➔ Enable connecting to existing clinical trials and
move genomic analyses into standard clinical practice

Challenges with FFPE?

- Difficulty of extracting samples
  - Deparrafinization & de-cross-linking of protein-DNA.
  - Physical size of the samples can be small
  - Yield

- Poor quality of extracted material due to:
  - Warm-ischemic time in operating room
  - Type of formalin used, how fixed, & how long (un-buffered vs. buffered)



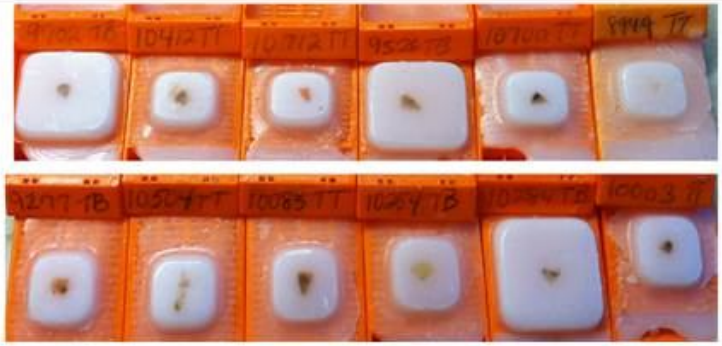
FFPE samples vary in size (TCGA samples)

FFPE Block Choices from Pilot Round #2

Group #1
(Large Tissues)

Group #2
(Medium Tissues)

Group #3
(Small Tissues)



Nationwide Children's Hospital
Biospecimen Core Resource

Samples from clinical study of drug resistance (Broad)

Small

Very small

Tiny

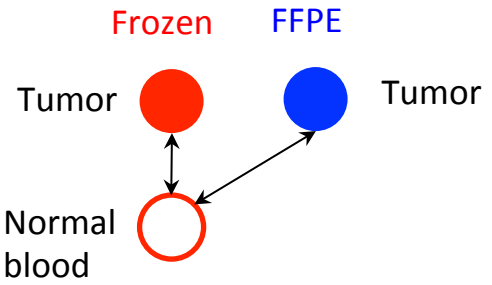
?

Where is it?

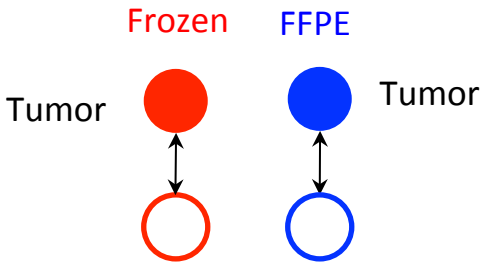
?

FFPE sample sets analyzed

- TCGA Prostate – "*trios*"
  - 4 FFPE Tumor samples + 4 Fresh Frozen Tumor/Normal pairs
  - Sequencing Coverage:
    - FFPE samples: 200x
    - Fresh Frozen pairs: 100x



- Breast Cancer – "*trios*"
  - 46 FFPE Tumor samples + 46 Fresh Frozen Tumor/Normal pairs
  - Source = FFPE Block, Mexico
  - Age of Fixed Block = 2008 – 2009 (plus a single 2010)



- Lung Cancer, NSCLC Adenocarcinoma – "*quartets*"
  - 17 FFPE Tumor/Normal sample pairs + 17 Fresh Frozen Tumor/Normal pairs
  - Source = FFPE Sections (15 microns, 9 per sample), Ontario, Canada
  - Age of Fixed Block = 2007 - 2010

Questions

1) Can we get high quality exome sequencing data from FFPE samples compared to frozen?

2) Can we detect mutations in FFPE samples? Are they artifacts?

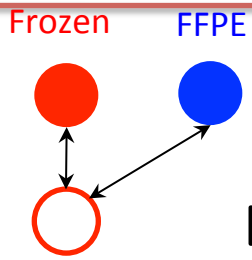
3) Can we detect copy-number changes?

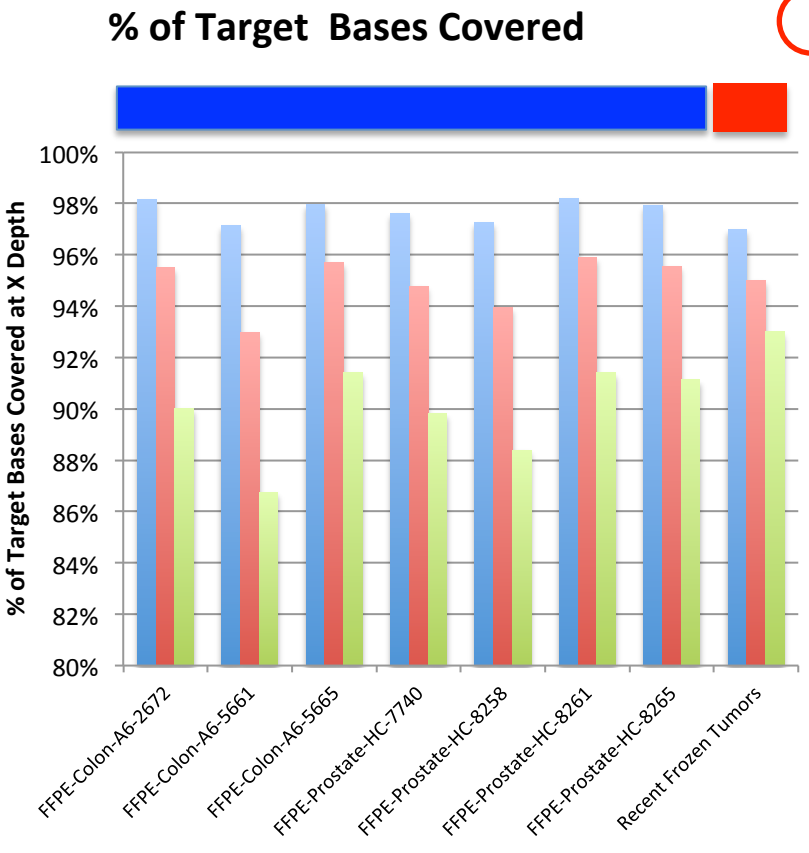
4) Are we finding the same mutations in FFPE vs. frozen?

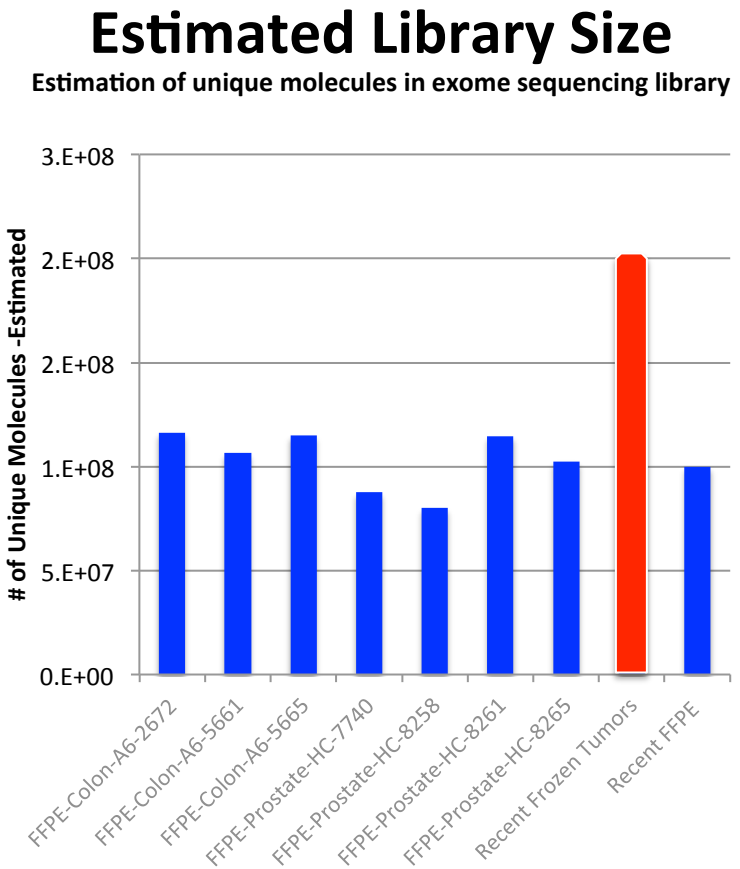
5) Can we perform cancer genome projects using FFPE samples?

6) Can we use clinical FFPE samples for clinical decision making?

(1) Are we getting similar library sizes from whole-exome sequencing?

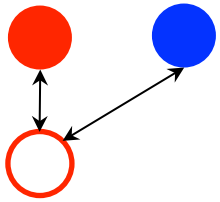






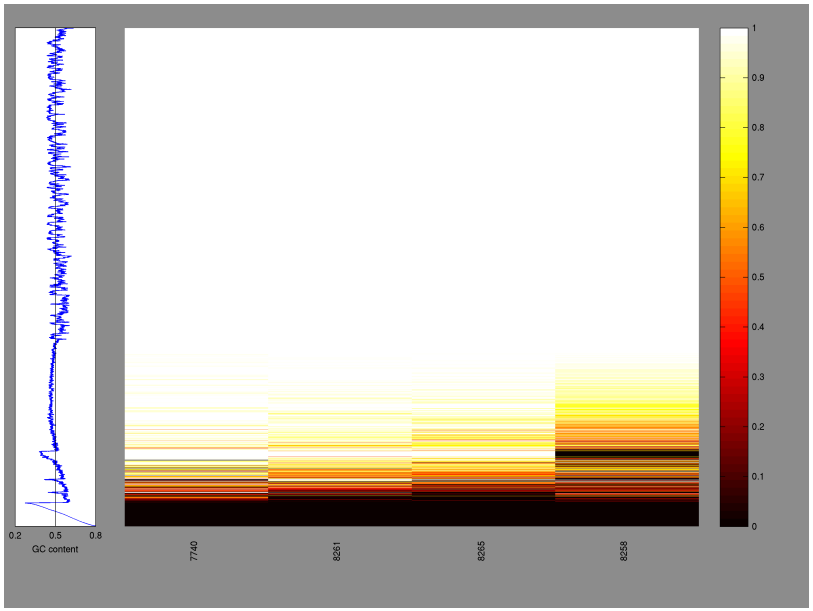
Carrie Sougnez

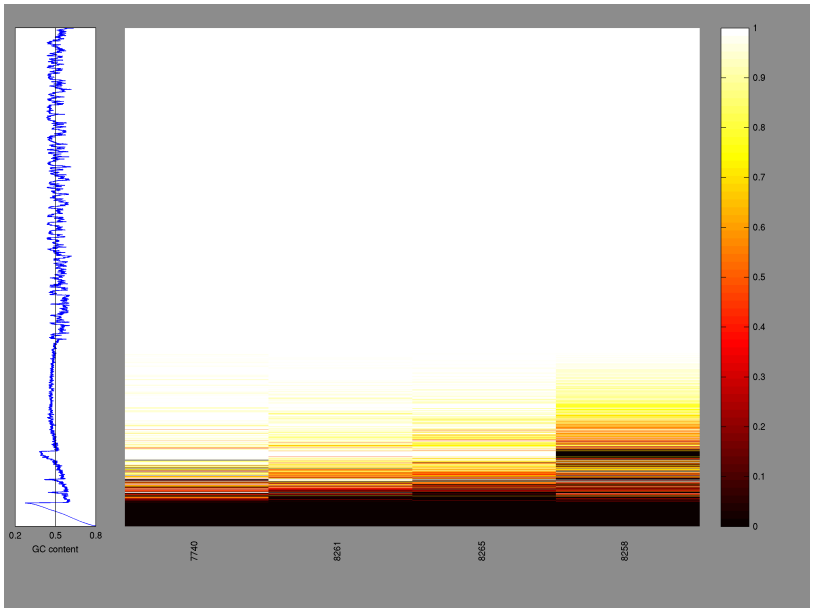
(1) Are we getting similar coverage? TCGA Prostate Cancer

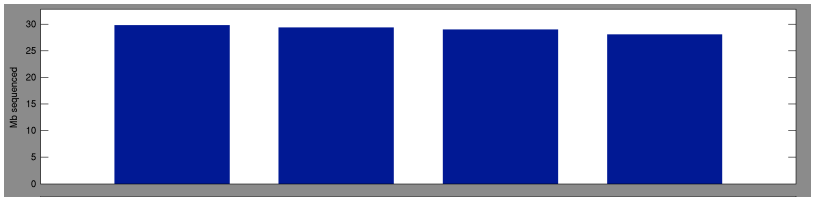


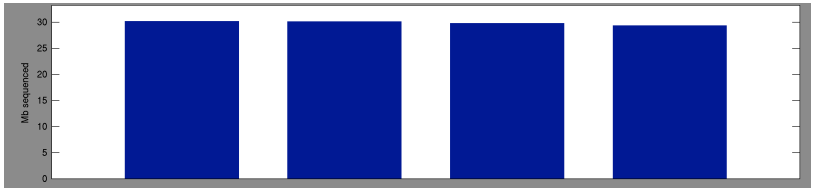
Frozen

FFPE



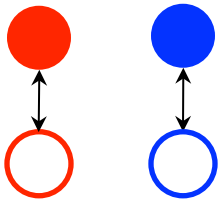




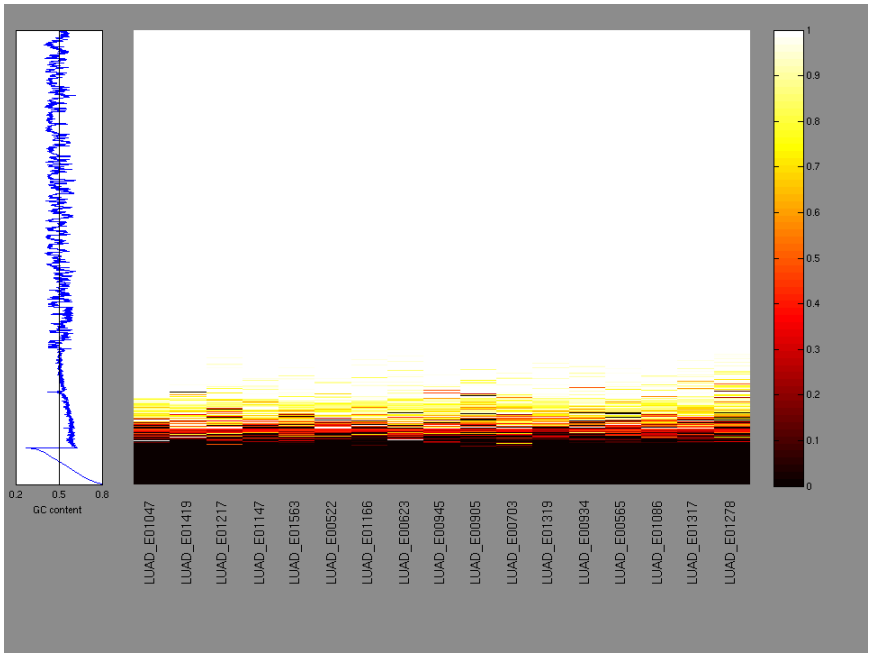


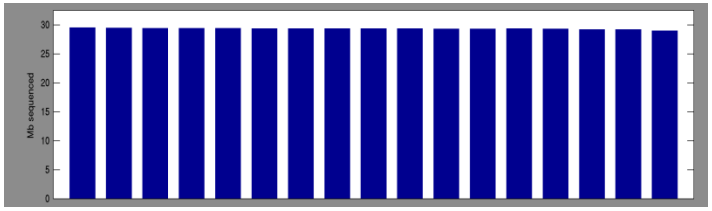
➔ Coverage (≥14 T/≥8 N) is roughly the same across all samples in Frozen and FFPE
~30Mb of covered bases

(1) Similar results in 17 Lung quartets: coverage statistics

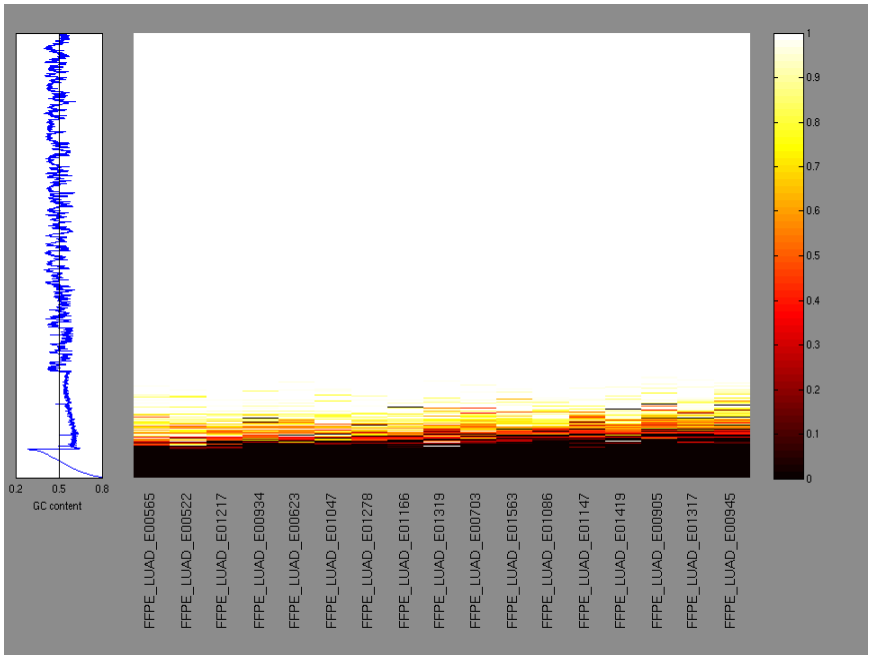


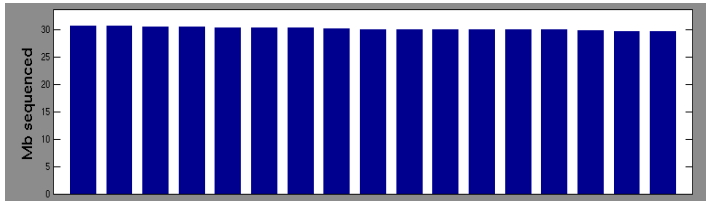
Frozen





FFPE





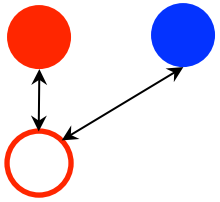
(2) Can we find mutations?
Total Count of mutations is similar

4 prostate

Frozen

type	count
Missense_Mutation	83
Nonsense_Mutation	8
Silent	41
Splice_Site	2
Translation_Start_Site	1
Total	135

Total territory: 130.49 MB



FFPE

type	count
Missense_Mutation	83
Nonsense_Mutation	6
Nonstop_Mutation	1
Silent	40
Splice_Site	6
Translation_Start_Site	1
Total	137

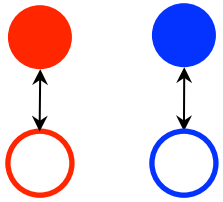
Total territory: 130.66 MB

17 lung

Frozen

type	count
De_novo_Start_OutOfFrame	2
Frame_Shift_Del	25
Frame_Shift_Ins	9
In_Frame_Del	11
In_Frame_Ins	2
Missense_Mutation	3651
Nonsense_Mutation	286
Nonstop_Mutation	7
Silent	1225
Splice_Site_DNP	4
Splice_Site_SNP	109
Start_Codon_Del	1
Total	5332

Total territory: 499.2 Mb



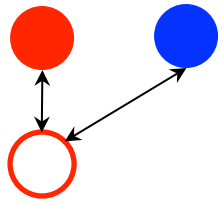
FFPE

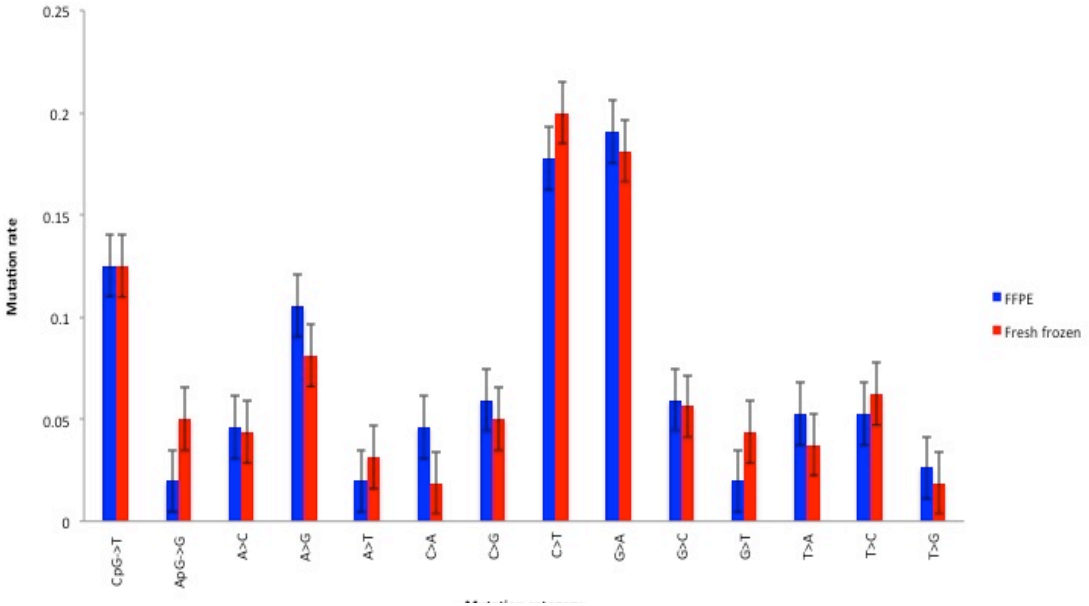
type	count
De_novo_Start_OutOfFrame	1
Frame_Shift_Del	17
Frame_Shift_Ins	4
In_Frame_Del	8
In_Frame_Ins	3
Missense_Mutation	3428
Nonsense_Mutation	270
Nonstop_Mutation	3
Silent	1152
Splice_Site_DNP	4
Splice_Site_SNP	122
Splice_Site_TNP	1
Total	5013

Total territory: 512.2 Mb

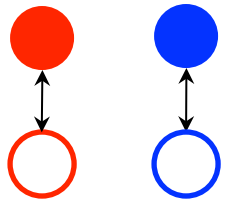
(2) Are the FFPE mutations swamped by artifacts? No! The mutations have the same spectra

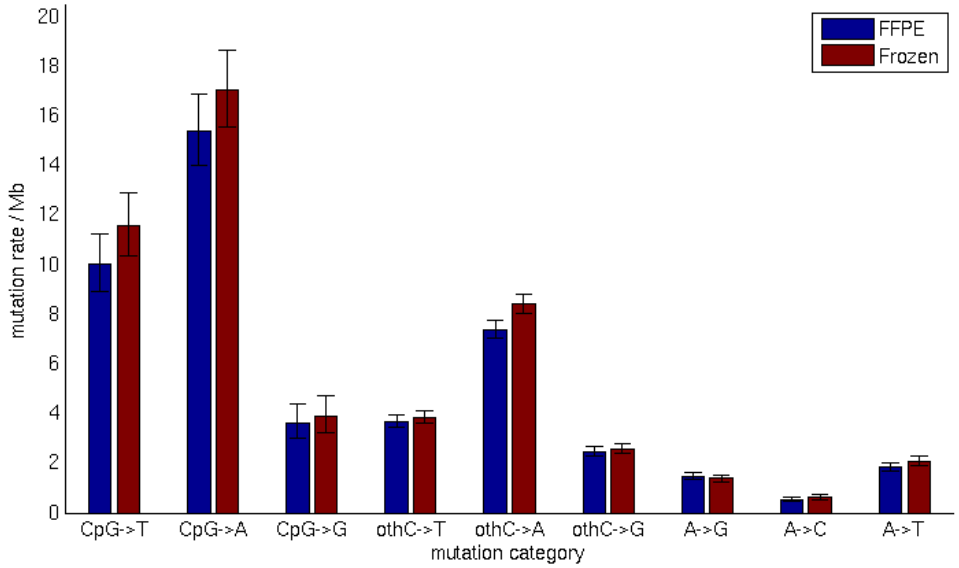
4 prostate



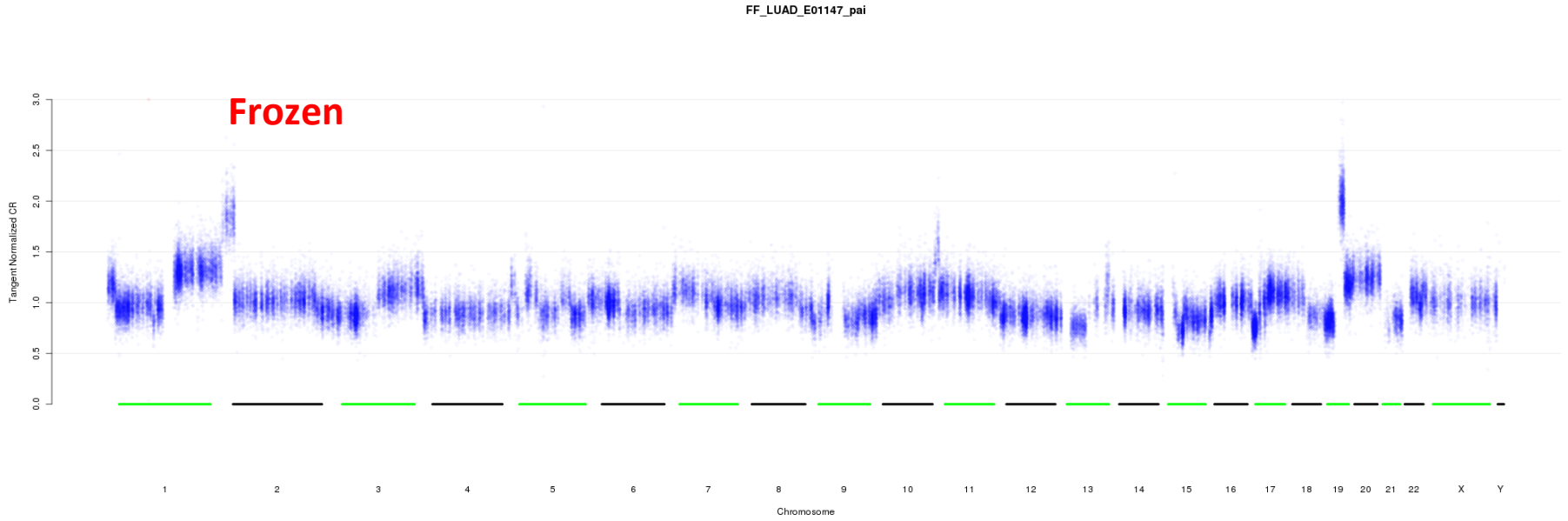


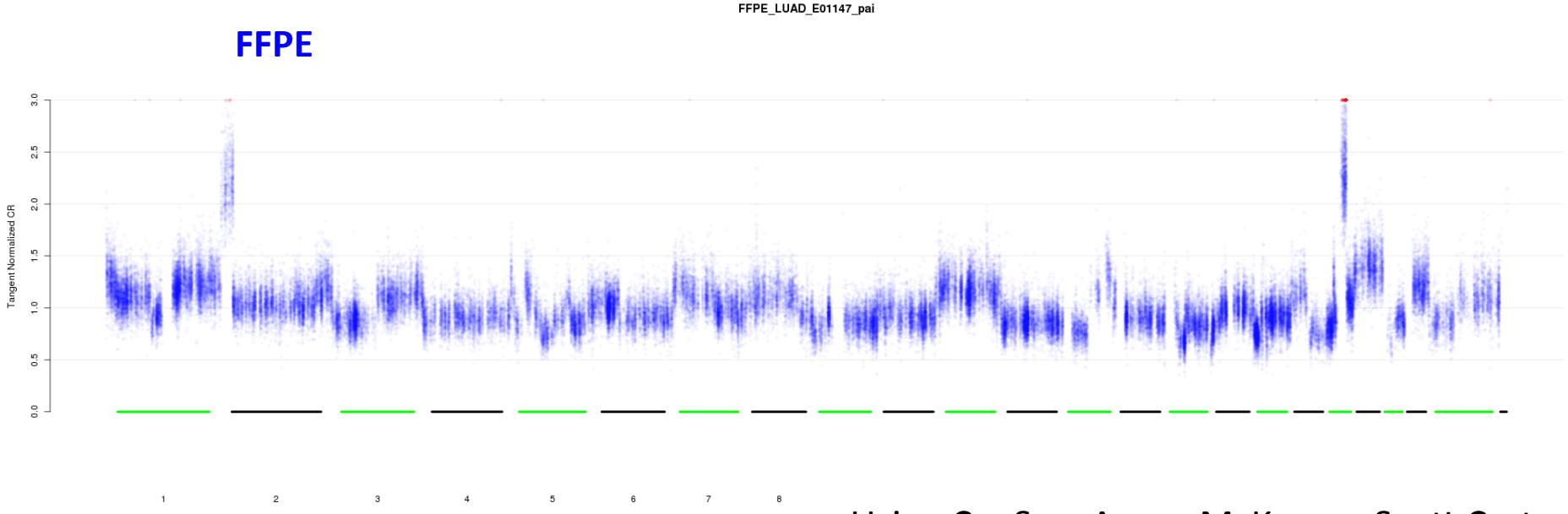
17 lung





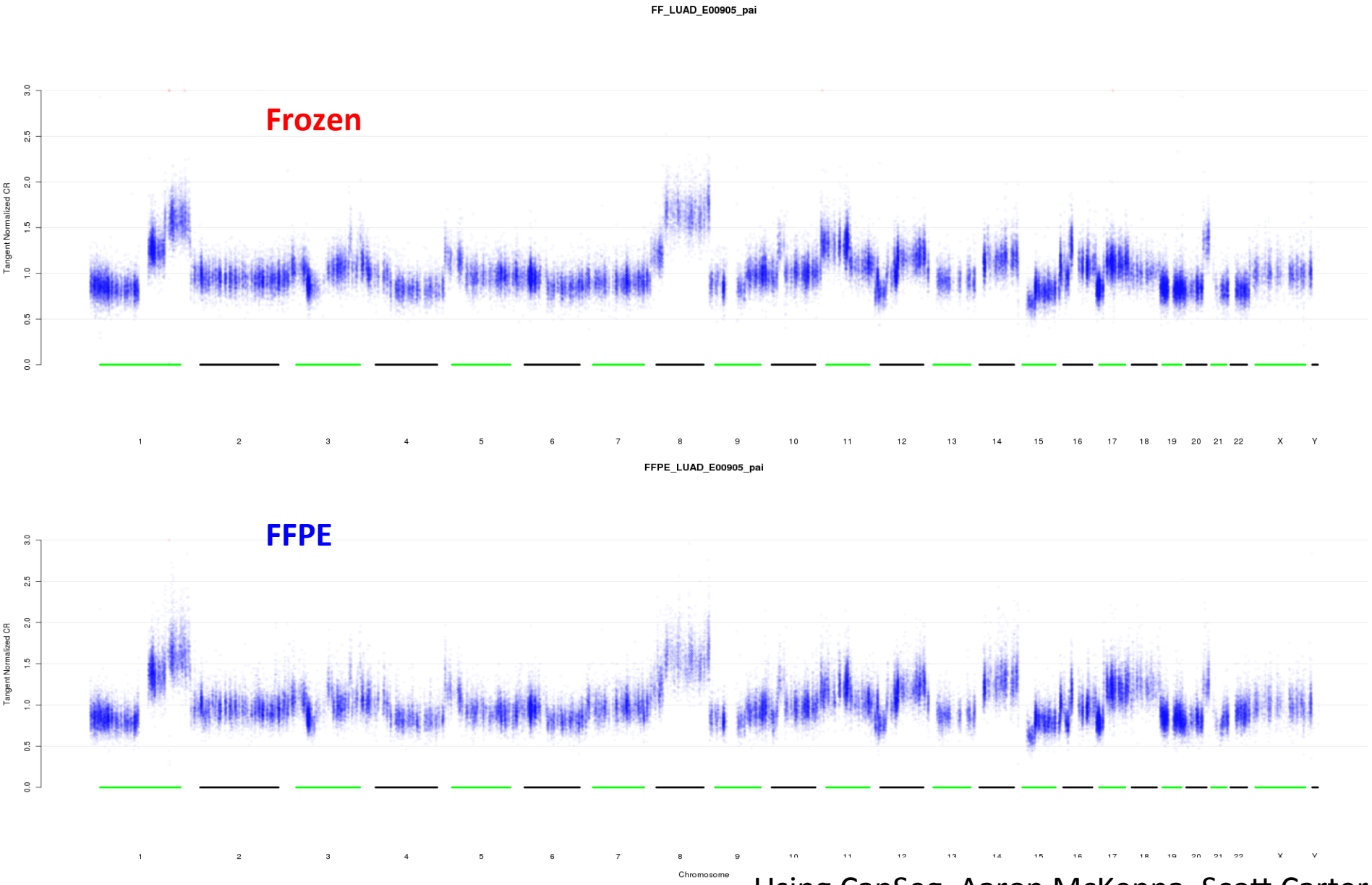
(3) Can we detect copy number changes? Example 1





Using CapSeg, Aaron McKenna, Scott Carter

(3) Can we detect copy number changes? Example 2

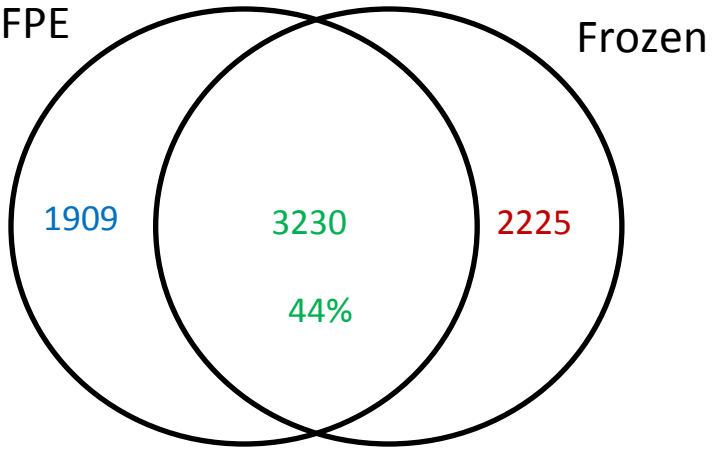


Using CapSeg, Aaron McKenna, Scott Carter

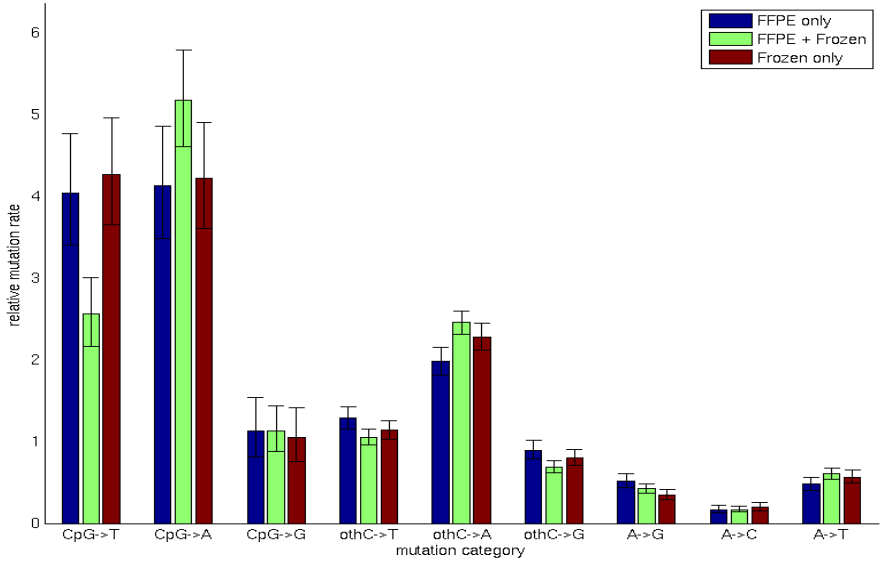
(4) Are we finding the same mutations in FFPE and frozen?

Overlap between FFPE and Frozen (17 lung)

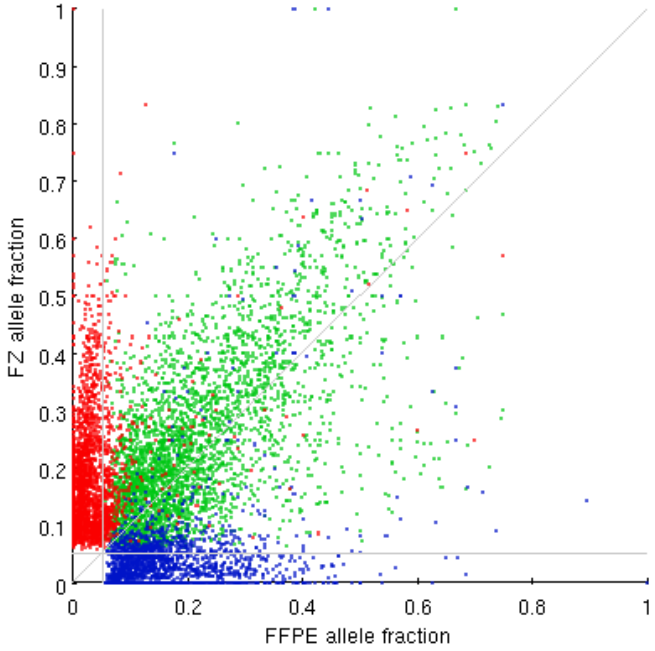
a



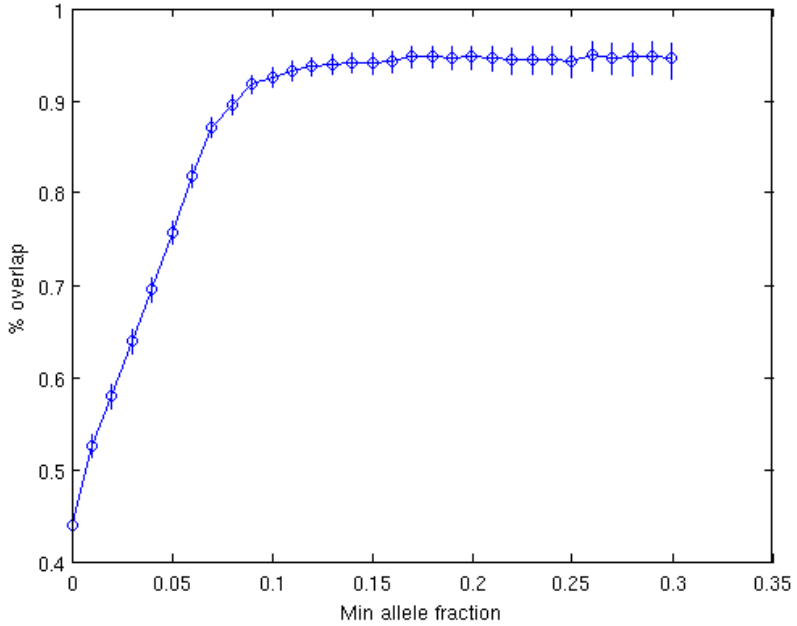
b



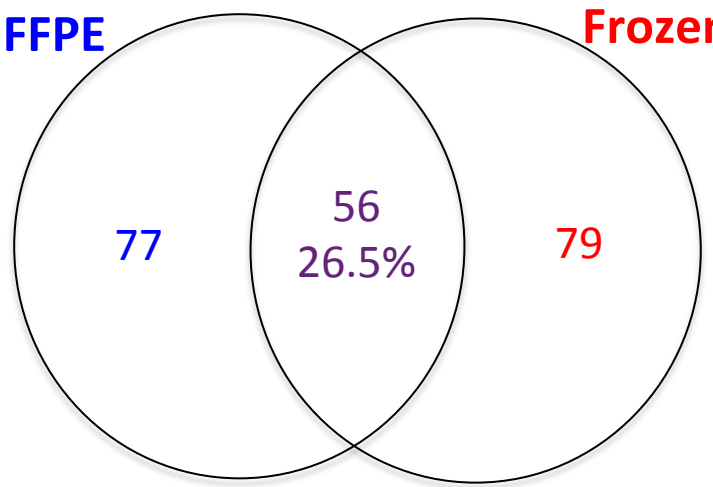
c

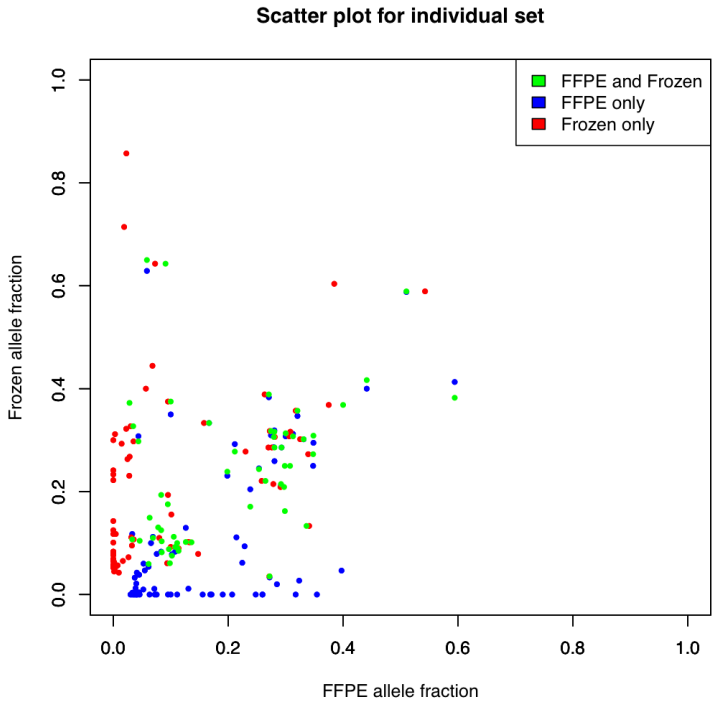


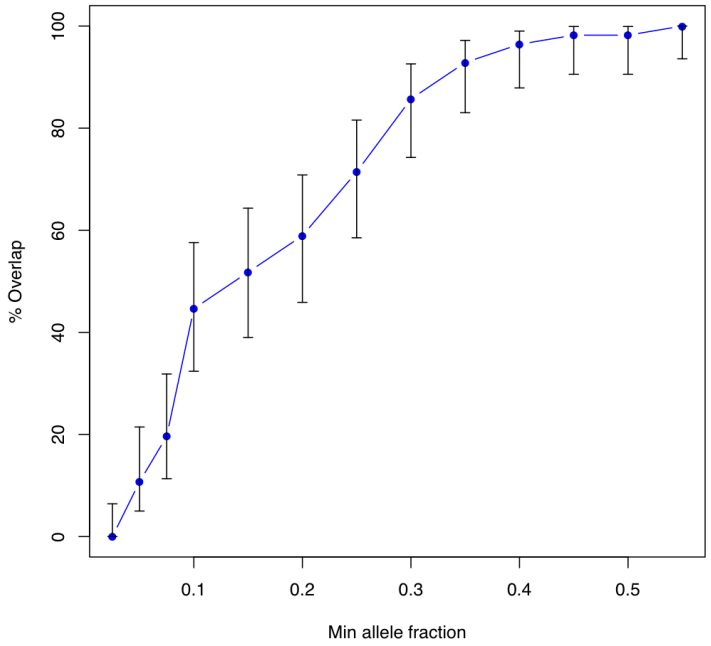
d



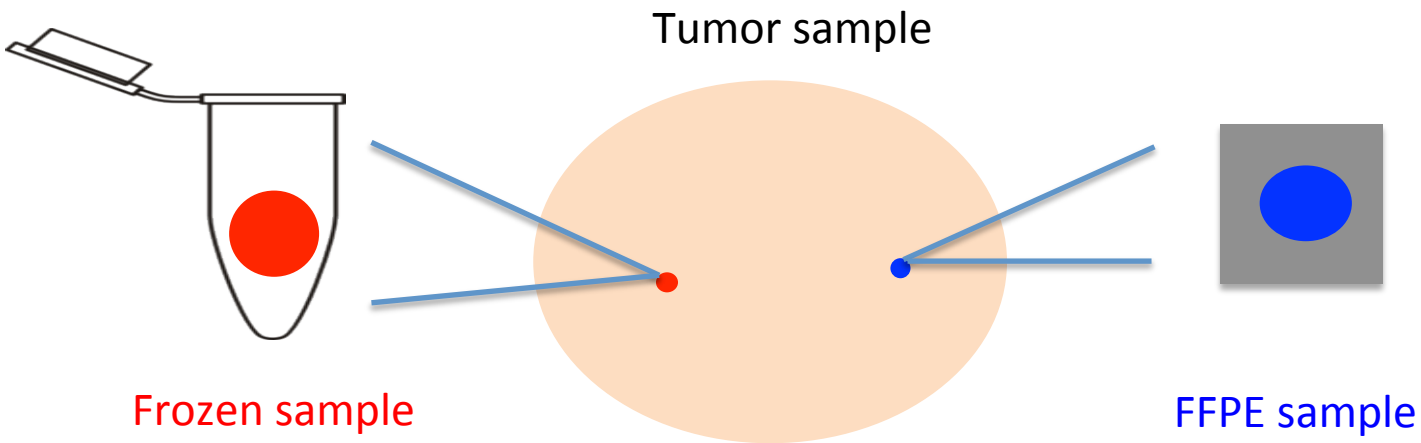
Overlap between FFPE and Frozen samples (4 prostate)







A fundamental observation: When comparing frozen to FFPE we are changing TWO variables at once

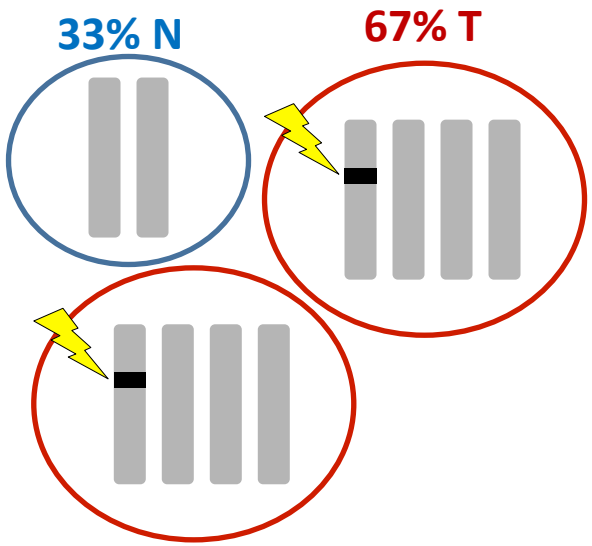


(1)Frozen vs FFPE
(2)Two different pieces of the tumor
-- Different in terms of tumor purity
-- Different with respect to sub-clonal composition

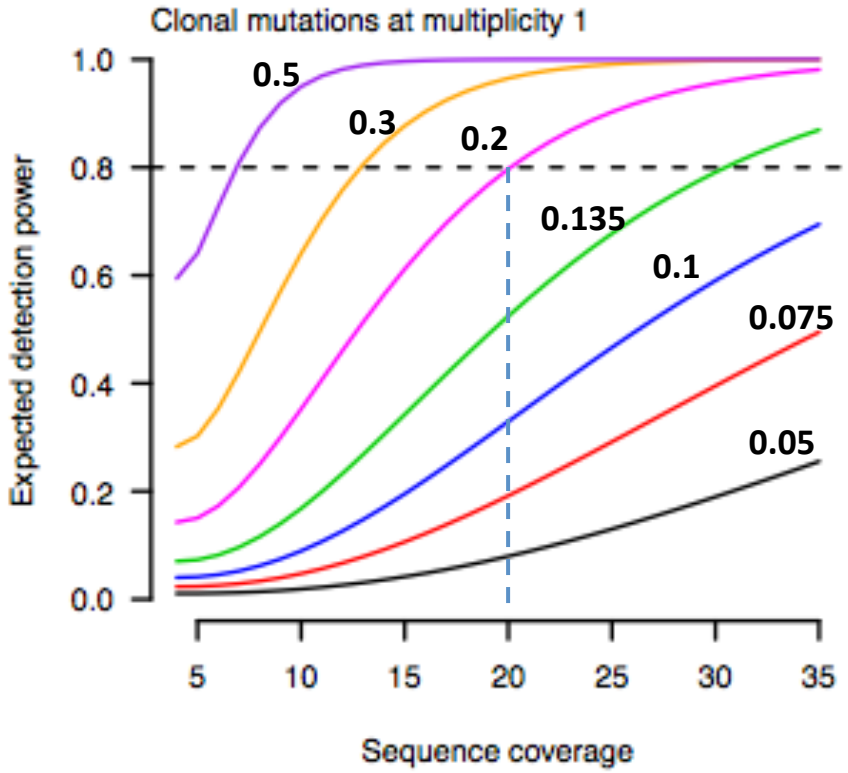
THIS AFFECTS ALL COMPARISONS BETWEEN
FFPE AND FROZEN SAMPLES (DNA, RNA, PROTEINS)

Sensitivity to detect (and even observe) a mutation – depends on coverage and allelic fraction

The ability to detect mutations depends on the **coverage** and **mutation allelic fraction** (the expected fraction of reads that support a mutation)



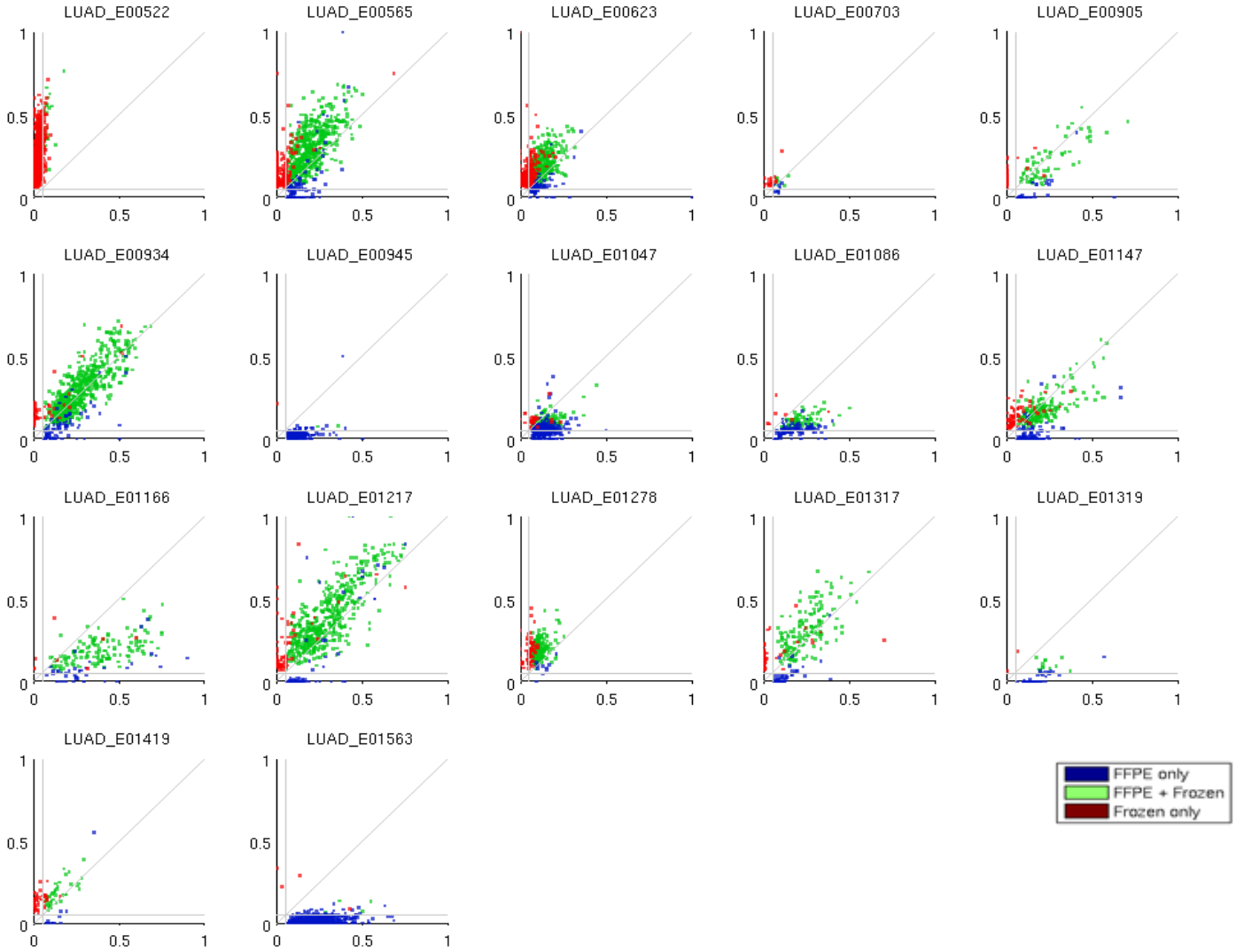
Purity = 67%
Absolute copy number in tumor = 4
Mutation multiplicity = 1
➔ Allelic fraction = 2/10 = 0.2



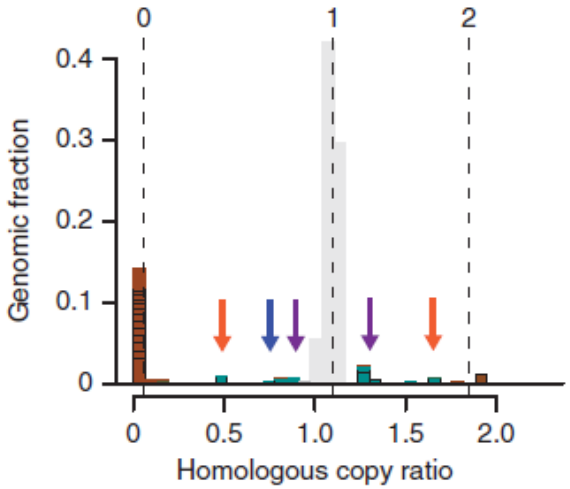
ABSOLUTE Carter et al. *Nat. Biotech.* (2012)

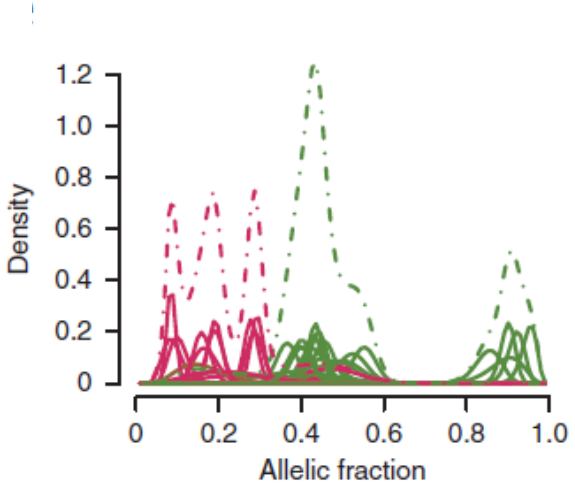
ABSOLUTE: SNP arrays / exome sequencing → purity, ploidy & abs. copy-number profile

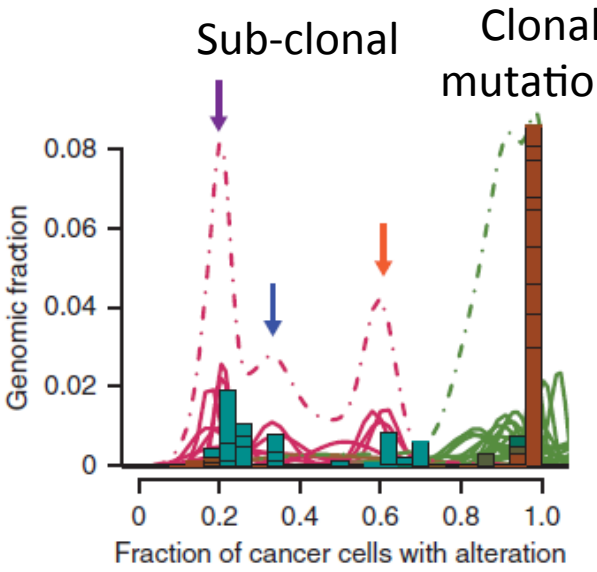
Allelic fraction in frozen and FFPE are different due to differences in purities (17 lung)



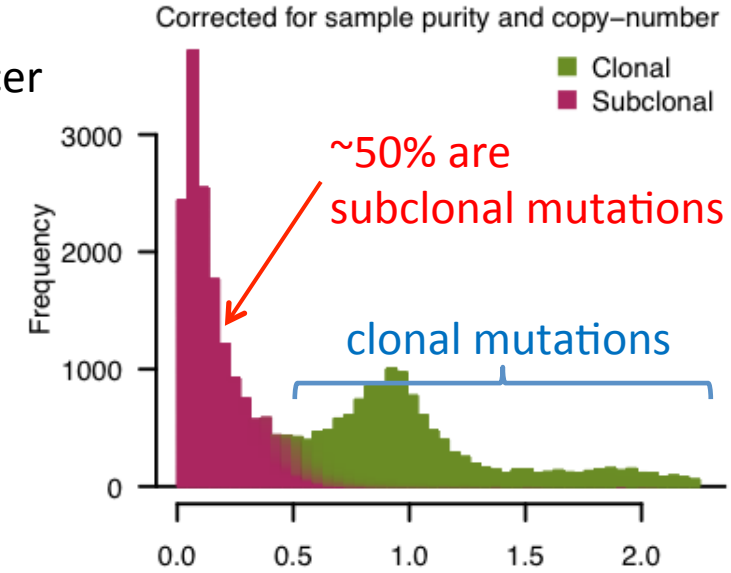
ABSOLUTE can distinguish between clonal & sub-clonal mutations







Ovarian cancer

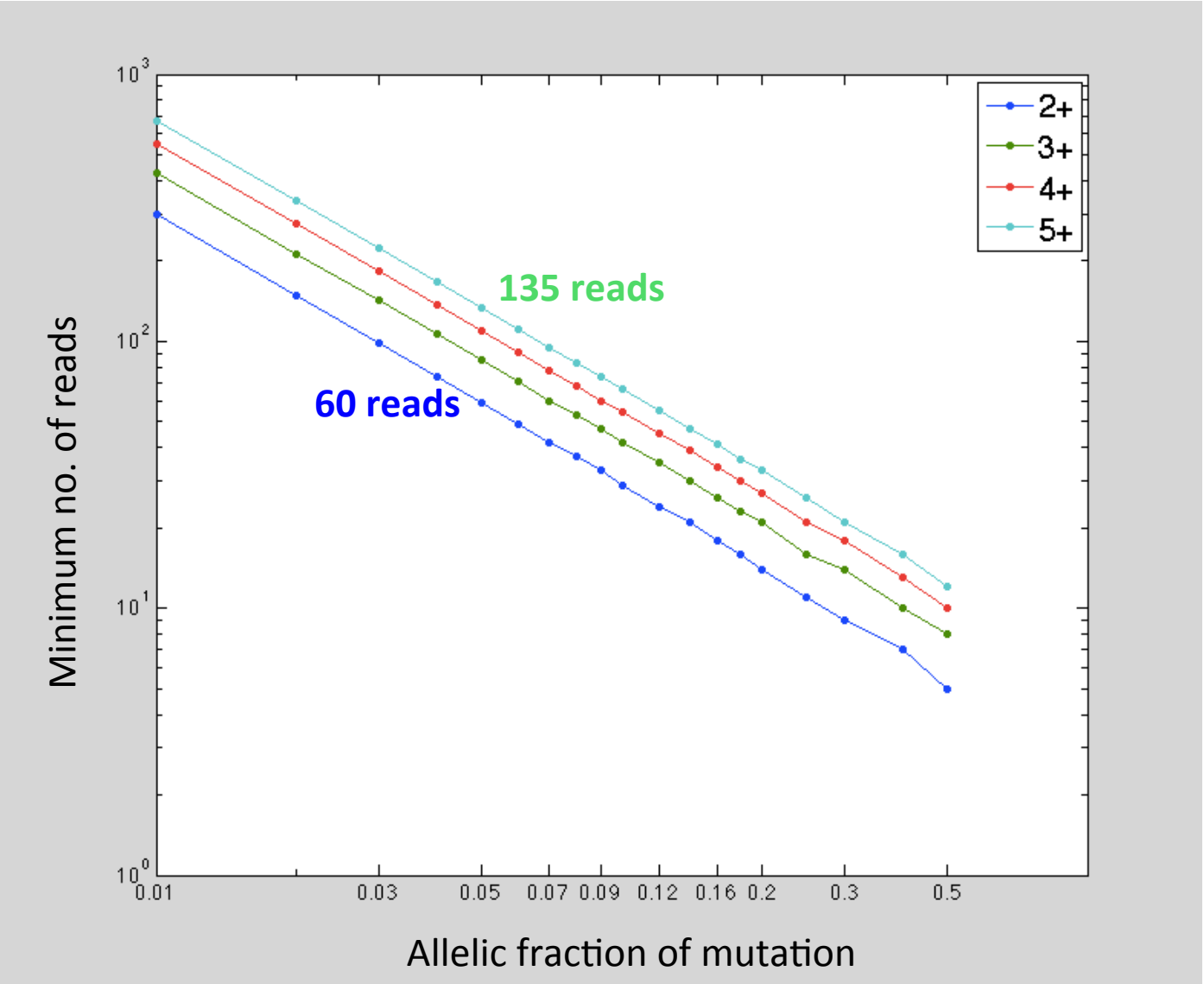


Carter et al. *Nat. Biotech.* (2012)

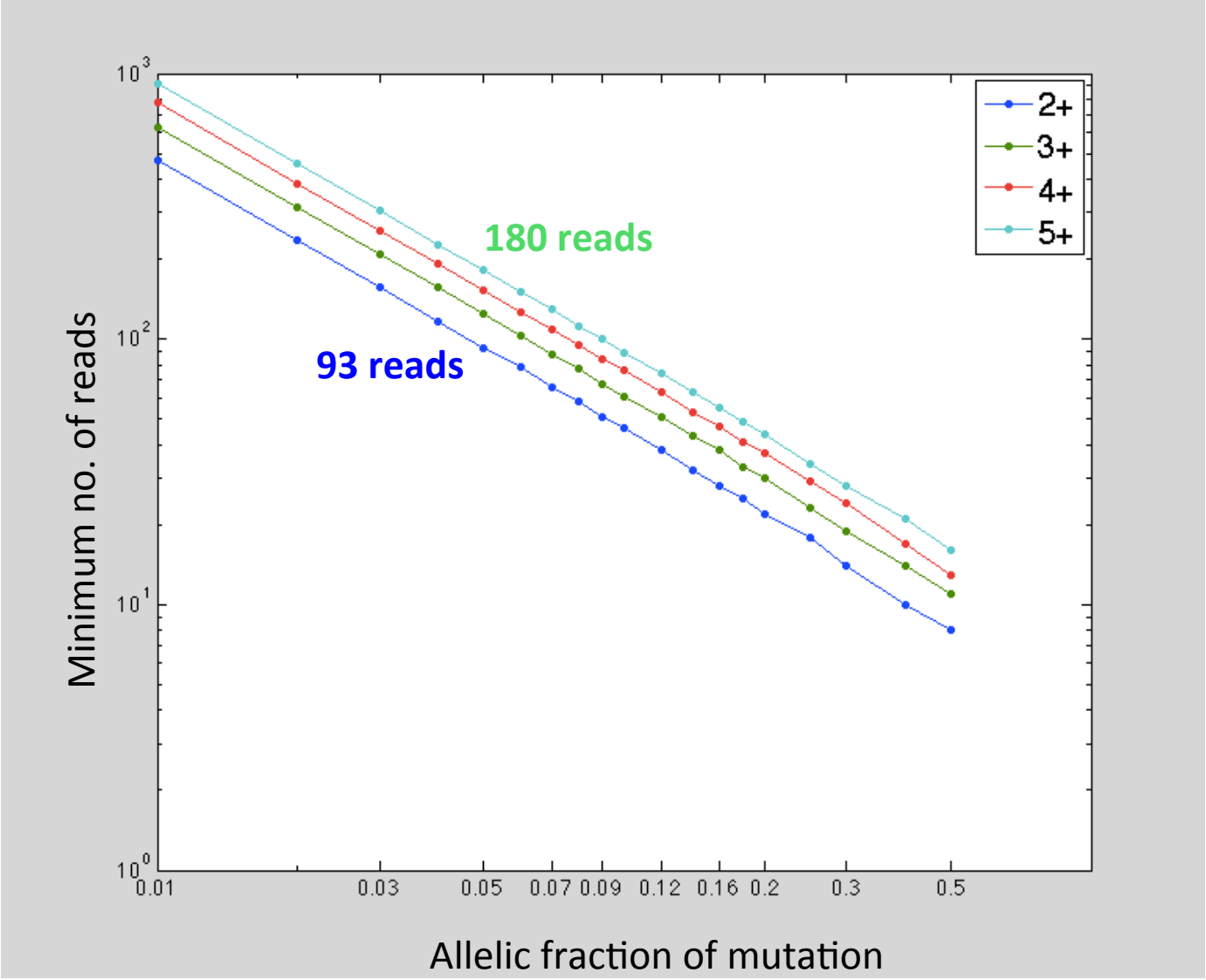
How should we compare the FFPE and frozen mutation sets?

1) We do not need to independently call the mutation in both FFPE and frozen. All we need is to validate the existence of the mutations found in FFPE in the frozen sample (i.e. call with a lower stringency since testing only a small number of mutation) → require 2+ reads
2) Correct for the different allelic fraction in the two samples due to different purity of FFPE and frozen → fit a line
3) Stratify sites based on the power to validate a mutation → 80%, 95%
4) Distinguish between clonal and sub-clonal mutations → use ABSOLUTE to assign mutation as clonal or sub-clonal

Minimum number of reads to have power of 80%

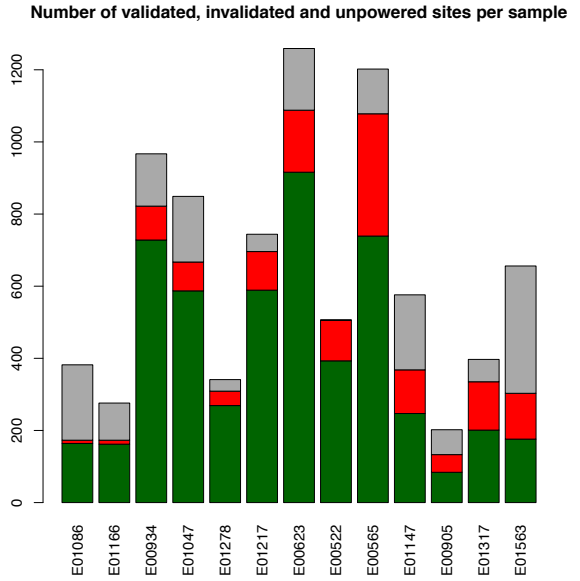


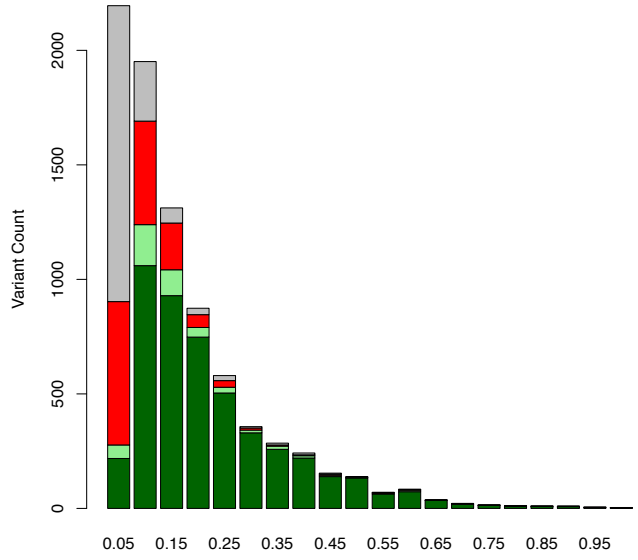
Minimum number of reads to have power of 95%



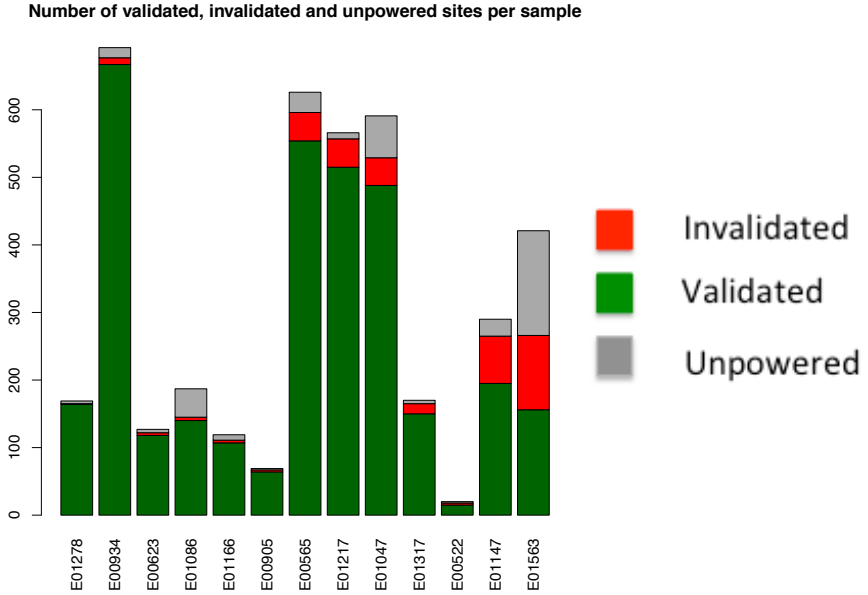
Validate = 2+ reads, AF corrected, power of 80%

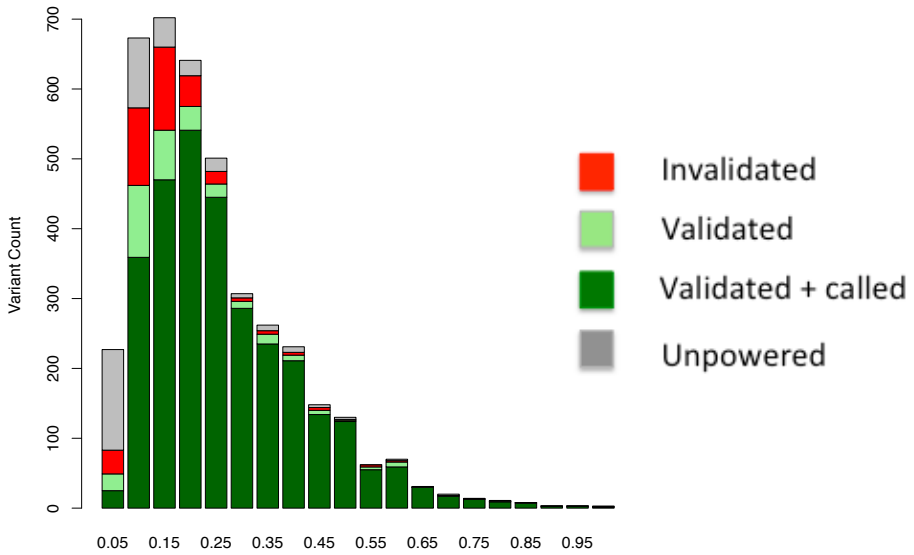
All





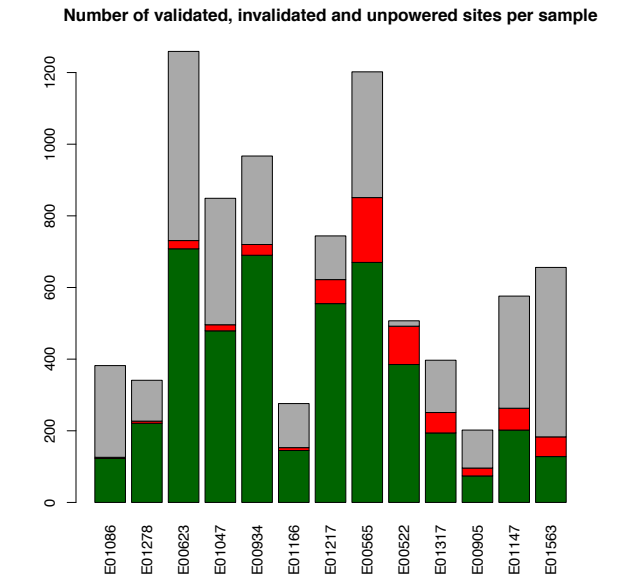
Clonal (based on ABSOLUTE on FFPE)

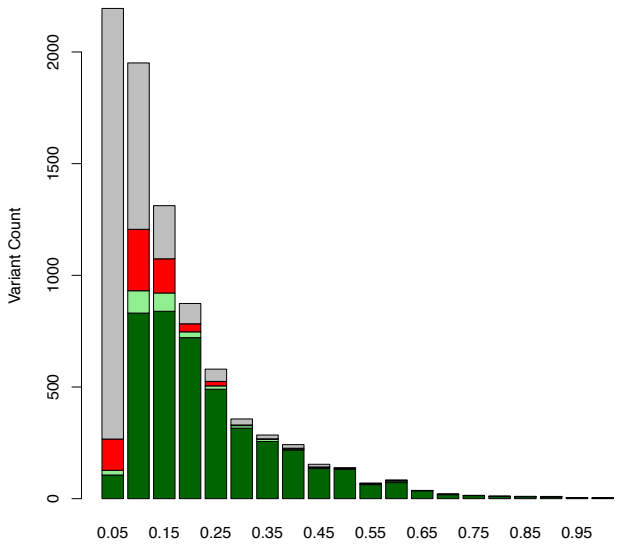




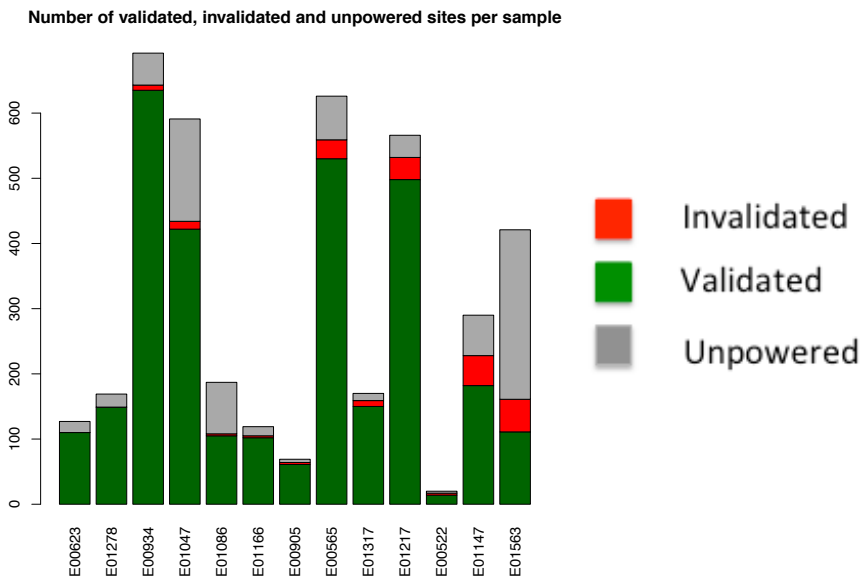
Validate = 2+ reads, AF corrected, power of 95%

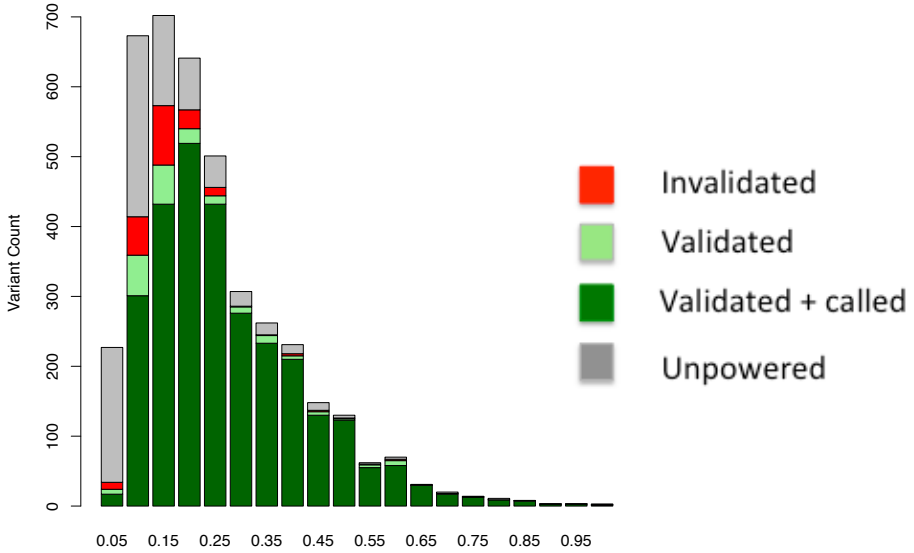
All





Clonal (based on ABSOLUTE on FFPE)



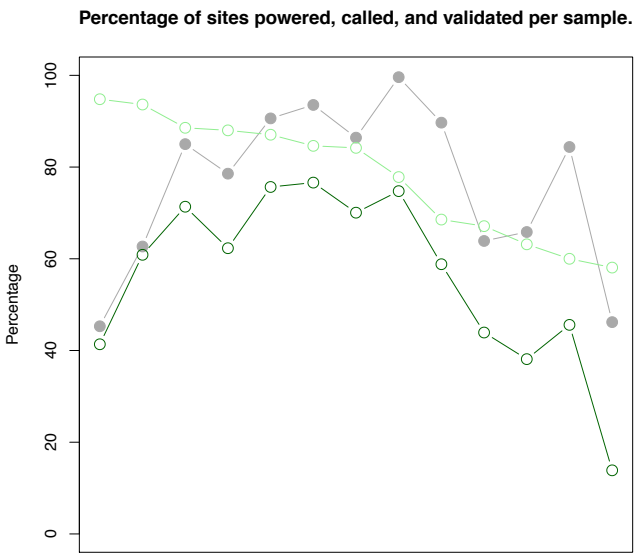


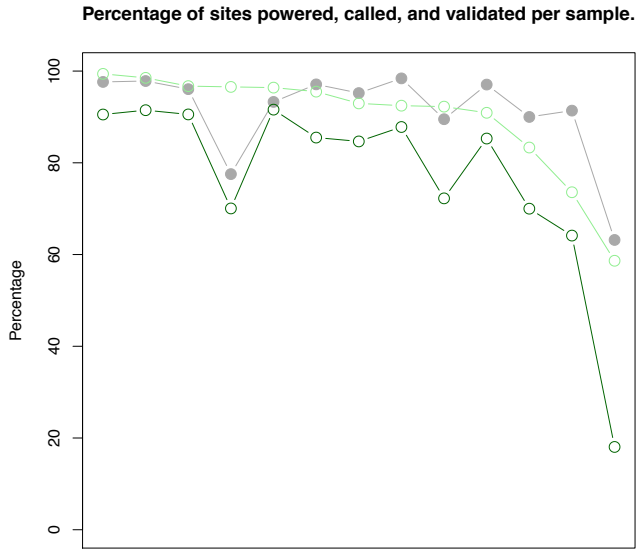
Validate = 2+ reads, AF corrected, power of 80% and 95%

All

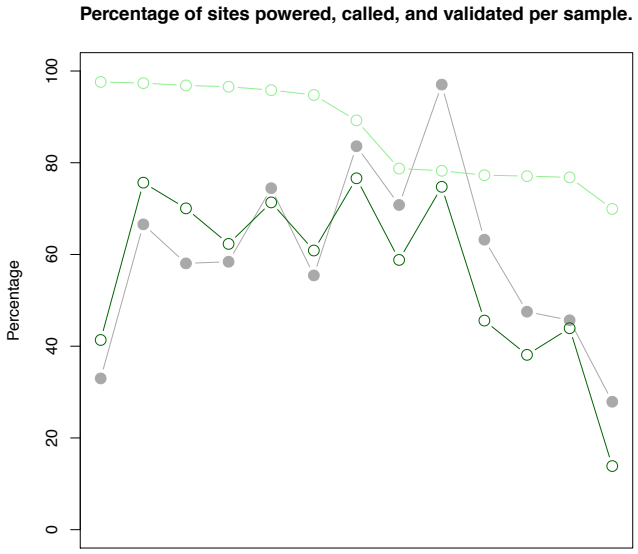
Clonal (based on ABSOLUTE on FFPE)

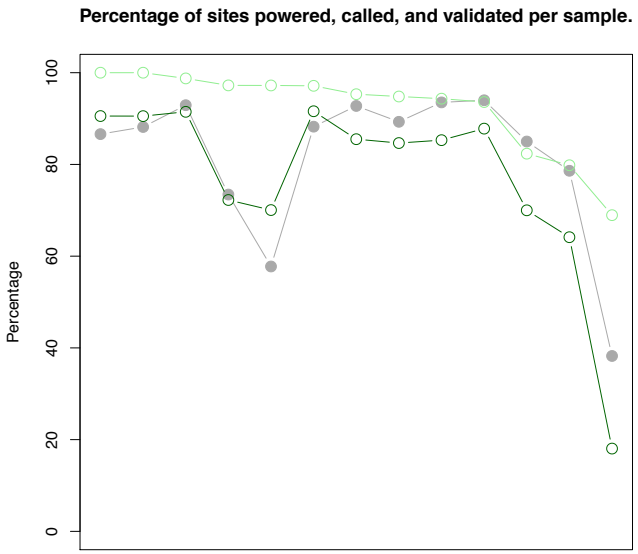
80%





95%





% powered

validated

Validated + called

(5) Can we perform cancer genome projects using FFPE samples?

MutSig: Significant genes (17 samples)

Frozen

rank	other rank	gene	description	N	n	npat	nsite	nsil	p	q
1	1	TP53	tumor protein p53	20661	7	7	7	0	8.77E-11	1.65E-06
2	2	KRAS	v-Ki-ras2 Kirsten rat sarcoma	11978	5	5	3	0	2.19E-09	0.000021
3	9	FAM5C	family with sequence similar	39593	7	4	7	1	5.36E-08	0.00034
4	30	CDH10	cadherin 10, type 2 (T2-cadhe	40987	6	4	6	0	9.93E-07	0.0047
5	4	OR6K2	olfactory receptor, family 6, :	16643	4	3	4	0	4.96E-06	0.019
6	13	CTNNA3	catenin (cadherin-associatec	46851	5	3	5	0	6.49E-06	0.02
7	14	NLGN1	neuroligin 1	42364	5	3	5	0	8.94E-06	0.022
8	1161	ERBB4	v-erb-a erythroblastic leuker	68595	6	4	6	0	9.19E-06	0.022
9	57	FAM135B	family with sequence similar	73049	6	6	6	2	0.000016	0.033
10	259	HIST1H1C	histone cluster 1, H1c	10982	3	3	3	0	0.000026	0.043
11	53	PTPRD	protein tyrosine phosphatas	99097	7	5	7	2	0.000026	0.043
12	193	ARHGAP15	Rho GTPase activating protei	25148	4	3	4	1	0.000031	0.043
13	570	ZNF479	zinc finger protein 479	27047	4	3	4	1	0.000035	0.043
14	16	ITGA4	integrin, alpha 4 (antigen CD	54557	5	4	5	1	0.000036	0.043
15	17	PSG9	pregnancy specific beta-1-gly	22185	4	3	4	1	0.000037	0.043
16	22	PTPRZ1	protein tyrosine phosphatas	119439	6	4	6	0	0.00004	0.043
17	43	OTUD6A	OTU domain containing 6A	10867	3	3	3	0	0.000041	0.043
18	66	A2BP1		24340	4	3	4	0	0.000041	0.043
19	128	POM121L12		15000	3	3	3	0	0.000045	0.045
20	1797	SYCP1	synaptonemal complex prote	46926	3	3	3	0	0.000051	0.048
21	1909	DPYS	dihydropyrimidinase	22576	3	2	3	0	0.000058	0.048
22	311	SETBP1	SET binding protein 1	73602	6	6	6	0	0.00006	0.048
23	2404	DNAH9	dynein, axonemal, heavy cha	225668	9	4	9	0	0.000061	0.048
24	2319	MMP16	matrix metallopeptidase 16 (	34340	4	3	4	0	0.000061	0.048
25	5	POTEH		9540	3	3	3	0	0.000063	0.048
26	1343	SORCS3	sortilin-related VPS10 domai	54215	5	3	5	1	0.000066	0.048
27	3	RYR2	ryanodine receptor 2 (cardiac	226445	9	5	9	1	0.00008	0.056
28	86	TGIF2LX	TGFB-induced factor homeob	12403	3	3	3	0	0.000086	0.058
29	68	FBXL7	F-box and leucine-rich repea	24561	4	3	4	0	0.000091	0.059
30	843	ABCB1	ATP-binding cassette, sub-fa	67167	5	4	5	0	0.000094	0.059
31	10	SLITRK1	SLIT and NTRK-like family, m	35615	4	2	4	0	0.00014	0.083
32	27	FAT3	FAT tumor suppressor homol	224971	9	4	9	2	0.00015	0.086
33	1492	SLC39A12	solute carrier family 39 (zinc	34153	4	3	4	1	0.00016	0.089
34	54	LRFN5	leucine rich repeat and fibro	36954	4	3	4	1	0.00017	0.097
35	29	ATP5G2	ATP synthase, H+ transportin	9231	2	1	2	0	0.00018	0.099

FFPE

rank	other rank	gene	description	N	n	npat	nsite	nsil	p	q
1	1	TP53	tumor protein p53	20882	8	8	8	0	<1.00e-11	<1.42e-07
2	2	KRAS	v-Ki-ras2 Kirsten rat sa	12019	6	6	2	0	1.51E-11	1.42E-07
3	27	RYR2	ryanodine receptor 2 (	240076	12	5	12	2	7.24E-08	0.00046
4	5	OR6K2	olfactory receptor, far	16643	5	4	5	0	1.84E-07	0.00087
5	25	POTEH		9554	4	3	4	0	4.76E-07	0.0018
6	102	STK11	serine/threonine kina	14299	4	4	4	0	1.41E-06	0.0044
7	60	SPTA1	spectrin, alpha, erythr	126956	8	6	8	1	1.90E-06	0.0051
8	207	OR8I2	olfactory receptor, far	15893	4	4	4	0	2.29E-06	0.0054
9	3	FAM5C	family with sequence	39593	5	3	5	1	5.98E-06	0.012
10	31	SLITRK1	SLIT and NTRK-like fan	35615	5	3	5	1	6.32E-06	0.012
11	39	OR4A15	olfactory receptor, far	17646	4	3	4	0	8.35E-06	0.014
12	144	AMPD1	adenosine monophos	39406	5	5	5	0	9.52E-06	0.015
13	6	CTNNA3	catenin (cadherin-ass	46852	5	4	5	1	0.000012	0.018
14	7	NLGN1	neuroligin 1	42364	5	3	5	0	0.000014	0.019
15	297	ZIC4	Zic family member 4	17245	4	4	4	1	0.000015	0.019
16	14	ITGA4	integrin, alpha 4 (anti	54587	4	3	4	1	0.000026	0.03
17	15	PSG9	pregnancy specific be	22185	4	3	4	1	0.000027	0.03
18	745	NDST3	N-deacetylase/N-sulfa	45448	4	3	4	0	0.000041	0.043
19	1016	HPSE2	heparanase 2	29031	4	3	4	0	0.000082	0.082
20	4773	BET1	blocked early in trans	6341	2	2	2	0	0.000088	0.082
21	83	SNAP25	synaptosomal-associa	13107	3	2	3	0	0.000091	0.082
22	16	PTPRZ1	protein tyrosine phos	120002	6	5	6	0	0.00011	0.092
23	1451	PRDM9	PR domain containing	46325	4	3	4	0	0.00011	0.092
24	165	CSMD3	CUB and Sushi multipl	194978	8	6	8	1	0.00015	0.11
25	81	BCL2L11	BCL2-like 11 (apoptosi	10353	2	1	2	0	0.00016	0.11
26	2630	PTCHD3	patched domain conta	39440	4	3	4	1	0.00016	0.11
27	32	FAT3	FAT tumor suppressor	229872	8	5	8	3	0.00017	0.12
28	105	GPR112	G protein-coupled rec	158668	6	6	6	2	0.00019	0.13
29	35	ATP5G2	ATP synthase, H+ tran	9491	2	1	2	0	0.00021	0.14
30	4	CDH10	cadherin 10, type 2 (T2	40987	4	3	4	1	0.00024	0.15
31	44	CLDN14	claudin 14	11143	2	2	2	0	0.00025	0.15
32	75	RGS5	regulator of G-protein	9622	2	2	2	0	0.00029	0.17
33	64	PSG6	pregnancy specific be	22083	3	2	3	0	0.00029	0.17
34	148	CDH4	cadherin 4, type 1, R-c	44958	4	4	4	0	0.0003	0.17
35	781	RGS22	regulator of G-protein	65769	4	4	4	0	0.00032	0.17

Orange background if within top 30 of other list

➔ Yes! Very similar MutSig lists

Old MutSig version

(6) Can we sequence clinical FFPE samples for clinical decision making? Yes!

17 lung samples

Gene	# in Frozen	# in FFPE
TP53	7	7
KRAS	4	5
EGFR	2	1
STK11	2	4
KEAP1	0	1
ATM	1	2
NF1	4	4

Conclusions

- Exome Sequencing of FFPE samples is robust – we can extract DNA, capture and sequence
- We can calculate overlap between FFPE and frozen samples controlling for relative coverage and adjust for different allelic fractions
- Mutation rates and categories are very similar
- Sub-clonal mutations contribute to the differences

➔ We can perform cancer genome project based on FFPE material

➔ We can use clinical FFPE samples for exome sequencing

- We are still analyzing more data in order to get reach final conclusions

Ongoing challenges

- WGS requires samples to be larger size range than exome (and sample prep more sensitive to changes that formalin fixation causes on DNA)
  - may not be suitable for samples that are highly degraded.
  - May need to optimize extraction steps to de-crosslink samples

- Low yield samples – small valuable specimens or micro-dissected samples

- Older blocks – may be very valuable but more variable due to storage conditions and older practices, such as use of unbuffered formalin (causes more DNA/RNA sample damage and cross-linking).

Acknowledgements

NCI/NHGRI

Kristin Ardlie

Petar Stojanov

Andrey Sivachenko

Scott Carter

Mike Lawrence

Carrie Sougnez

Daniel Auclair

Marcin Imilienski

Kristian Cibulskis

Stacey Gabriel

Matthew Meyerson

Todd Golub

Eric Lander

Broad

Biological Sample Platform

Genetic Analysis Platform

Sequencing Platform

TCGA

Kenna Shaw

Brad Ozenberger

NCH BCR

THE END