Next generation sequencing on FFPE tumor tissues for lung and colorectal carcinoma diagnostics and research

Sakari Knuutila

Tumor diagnostics and research

- **Why NGS or targeted deep sequencing?**
- Why not PCR or Sanger sequencing?

- **Why formalin fixed paraffin embedded tissue?**
- Why not fresh frozen tissues?

Simple because

- **NGS** makes it possible to analyze all kind of mutations (gene fusions and copy numbers included) of numerous genes simultaneously in one experiment
- **FFPE** blocks are used for basic histopathology diagnoses all over in standard pathology laboratories and in every laboratory we can find massive stores including clinical data

Aims

- To check feasibility of applying NGS to FFPE samples of lung adenocarcinoma and colorectal carcinoma.
- To compare mutations detected by prevalent methods & NGS.
- To mine novel clinically and biologically relevant genes
  - TK-genes
  - miRNA
  - Other *EGFR* and *ALK* signaling pathway genes

”Protocol 192”: Our patients

- Total 99 patients:
  - Lung adenocarcinoma (non-smokers) : 81 (*EGFR, KRAS*, *BRAF and ALK* status, and treatment result and survival known)
  - Colorectal carcinoma 18 (*KRAS, BRAF*, and *EGFR* status, and treatment result and survival known)

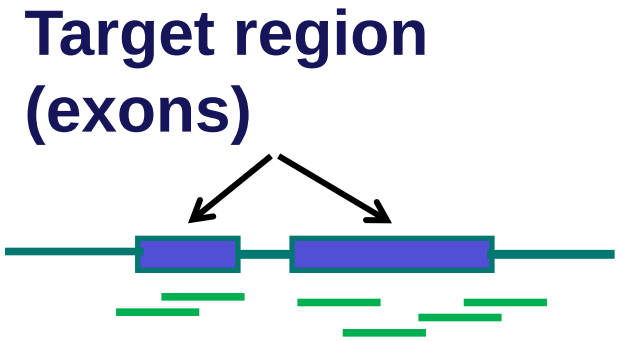
In total - 192 genes selected for targeted deep sequencing (TDS)

Basic steps for TDS

- Target gene selection (192 genes) and designing of baits
- Target capture and amplification of all exons of the 192 genes (intron for *ALK* gene)
- Sequencing
- Data analysis

TDS:Target capture

- Baits Designed by e-array (Agilent)
- Target region: ~ 1 MB
- Target regions: 2676
- Genes:192
  - *EGFR* pathways -related
  - Lung cancer -related
  - miRNA



- DNA
- Fragmentation
- Ligation with paired-end adaptors: Adaptor-ligated libraries
- Hybridization with biotinylated oligo RNA baits
- Enrichment with streptavidin congugated magnetic beads
- Further amplification, index taging
- & Illumina sequencing

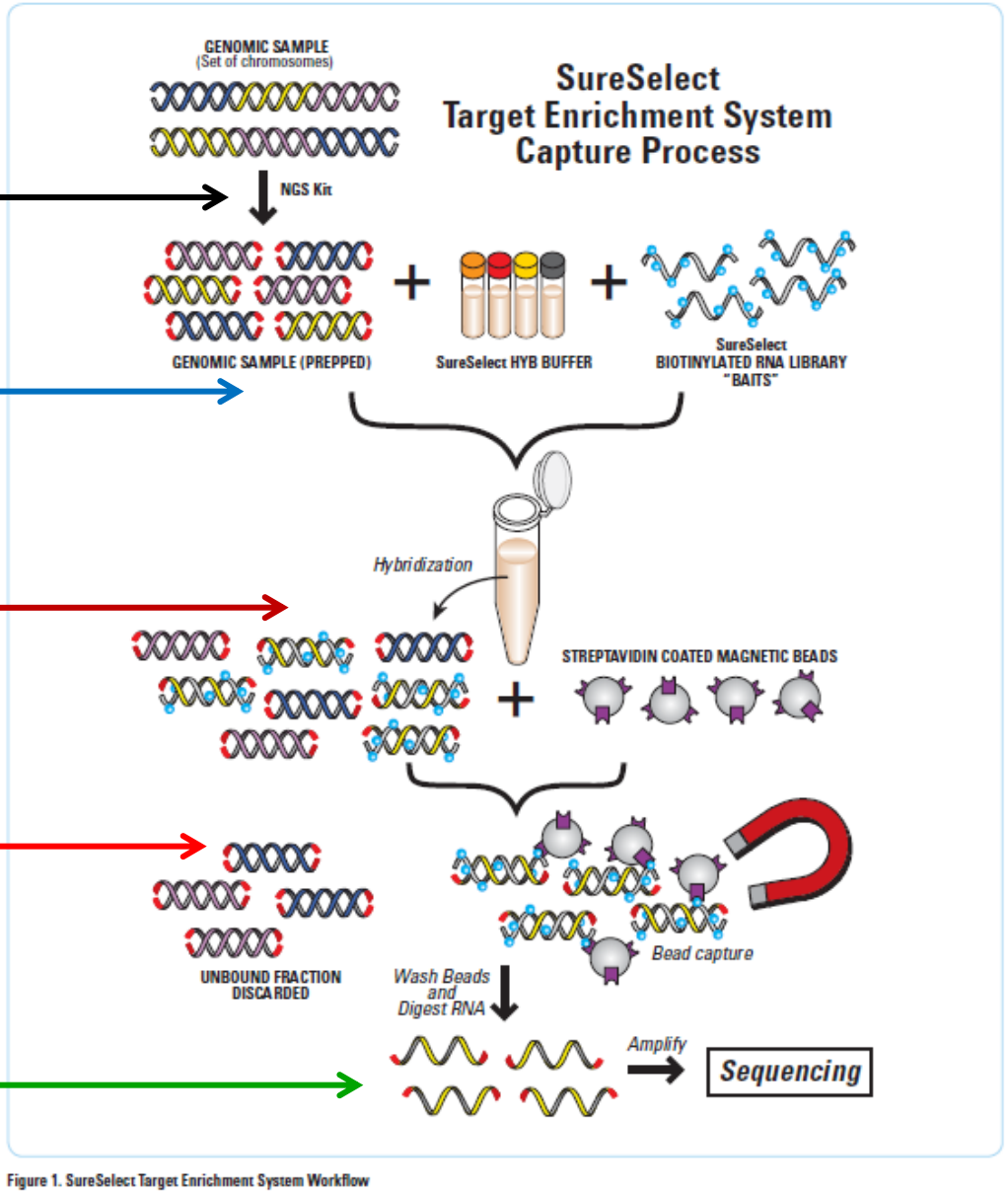


Figure 1. SureSelect Target Enrichment System Workflow

DNA extraction from FFPE specimens

- **QIAamp DNA Mini Kit:** according to the manufacturer´s instructions for genomic DNA purification from paraffin-embedded tissue material with following modifications: xylene was added twice and 100% ethanol was added three times for **careful paraffin removal**, samples were incubated at 90ºC for 1 h after overnight cell lysis, Buffer AL was added normally but an incubation at 70ºC for 10 min was not performed, and in the last step of the protocol, QIAamp Mini spin column loaded with water was incubated at room temperature for 5 min before first centrifugation, and at 50ºC for 5 min before the final centrifugation to increase the DNA yield. DNA was eluted in 50 µl of purified water. DNA **concentration** was measured **by Qubit® fluorometer** (Life Technologies).

Data analysis

- Data obtained from sequencing was processed with a **variant-calling pipeline** developed at the Finnish Institute of Molecular Medicine. Sequence reads were filtered for quality, paired-end reads aligned to **the reference genome with the Burrows-Wheeler Aligner** and **duplicate fragments removed by rmdup algorithm**. For **variant calling, SAMtools' pileup was utilized**. **SNV** calling **and read end anomaly** calling were based on **FIMM's** own developed **algorithm**. Detection of **indels** was performed with the **Pindel,** and **Circos** was used for visualization of anomalously mapping paired-end reads. Results were **visualized** in the **Integrative Genomic Viewer**.

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

- sakari.knuutila@helsinki.fi

Our Genes and miRNAs for 192 Protocol

Diagnostic validation

EGFR

KRAS

BRAF

ALK/ELM4 -fusion

ABL1, ABL2, AKT1, AKT2, AKT3, ALK, ARAF, AREG, AXL, BAD, BRAF, BTC, CAMK2A, CAMK2B, CAMK2D, CAMK2G, CBL, CBLB, CBLC, CCND1, CDKN1A, CDKN1B, CDKN2A, CRK, CRKL, CSF1R, CTNNB1, DDR1, DDR2, EGF, EGFR, EIF4EBP1, ELK1, EML4, EPHA1, EPHA2, EPHA3, EPHA4, EPHA5, EPHA6, EPHA7, EPHA8, EPHB1, EPHB2, EPHB3, EPHB4, EPHB6, EPHX1, ERBB2 , ERBB3 , ERBB4 , EREG , FGFR1, FGFR2, FGFR3, FGFR4, FLT1, FLT3, FLT4, GAB1 , GRB2, GSK3B, HBEGF, HMGA1, HMGA2, HRAS, IGF1, IGF1R, INSR, INSRR, JUN, KDR, KIT, KRAS, LTK, MAP2K1, MAP2K2, MAP2K4, MAP2K7, MAPK1, MAPK10, MAPK3, MAPK8 , MAPK9, MERTK, MET

MIR125A, MIR126, MIR135B, MIR137, MIR143, MIR152, MIR155, MIR15A, MIR16-1, MIR17HG, MIR181A1, MIR181A2, MIR181D, MIR182, MIR18A, MIR19A, MIR200C , MIR205, MIR206, MIR20A, MIR21, MIR210, MIR221, MIR222, MIR26A1, MIR29A, MIR29C, MIR335, MIR34A, MIR34B, MIR34C, MIR372, MIR378, MIR451, MIR7-1, MIR92A1, MIR98, MIRLET7A1, MIRLET7A2, MIRLET7A3, MIRLET7B, MIRLET7C, MIRLET7D, MIRLET7E, MIRLET7F1, MIRLET7F2, MIRLET7G, MIRLET7I

MST1R, MTOR, MUSK, MYC, NCK1, NCK2, NFE2L2, NRAS, NRG1, NRG2, NRG3, NRG4, PAK1, PAK2, PAK3, PAK4, PAK6, PAK7, PDGFRA, PDGFRB, PDGFRL, PIK3CA, PIK3CB, PIK3CD, PIK3CG, PIK3R1, PIK3R2, PIK3R3, PIK3R5, PLCG1, PLCG2, PRKCA, PRKCB, PRKCG, PTEN, PTK2, PTK7, RAF1, REL, RET, RPS6KB1, RPS6KB2, SHC1, SHC2 , SHC3, SHC4, SOS1, SOS2, SRC, STAT5A, STAT5B, STK11, TGFA , TGFB1, TP53, TRAF6, TYRO3, VEGFA, VEGFB

Red: targeted treratment available (far from complete list)

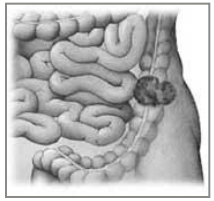
EGFR Mutation Kit – ARMS based

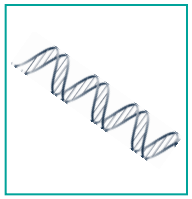
- TheraScreen EGFR28 Mutation Test Kit from DxS.
  - Link to DxS Website
- This kit is an ARMS based method which analyses 28 mutations within the EGFR gene:
  - 19 Deletions in exon 19
  - L858R
  - L861Q
  - G719X (G719S/G719A /G719C)
  - S768I
  - 3 insertions in exon 20

Newton *et al.*, 1989
Whitcombe *et al.*, 1999

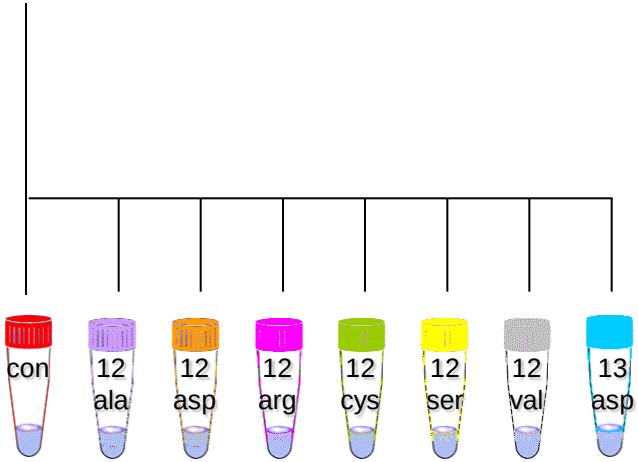
ARMS Scorpions Workflow

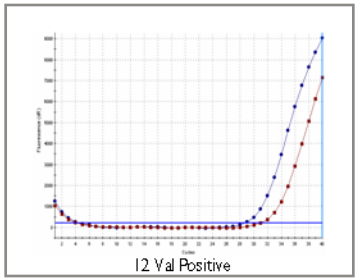
DNA is extracted from formalin fixed paraffin embedded tissue using standard methodology





Each DNA sample is added to 8 separate reaction tubes and placed onto a real-time PCR instrument





The analysis is completed by comparing the mutant reactions to the normal

The PCR reaction takes around 90 minutes

Concordance of the results of *EGFR, KRAS* and *BRAF* mutations by targeted deep sequencing (TDS) and PCR in lung adenocarcinoma (LAC) and metastasizing colorectal carcinoma (CRC)

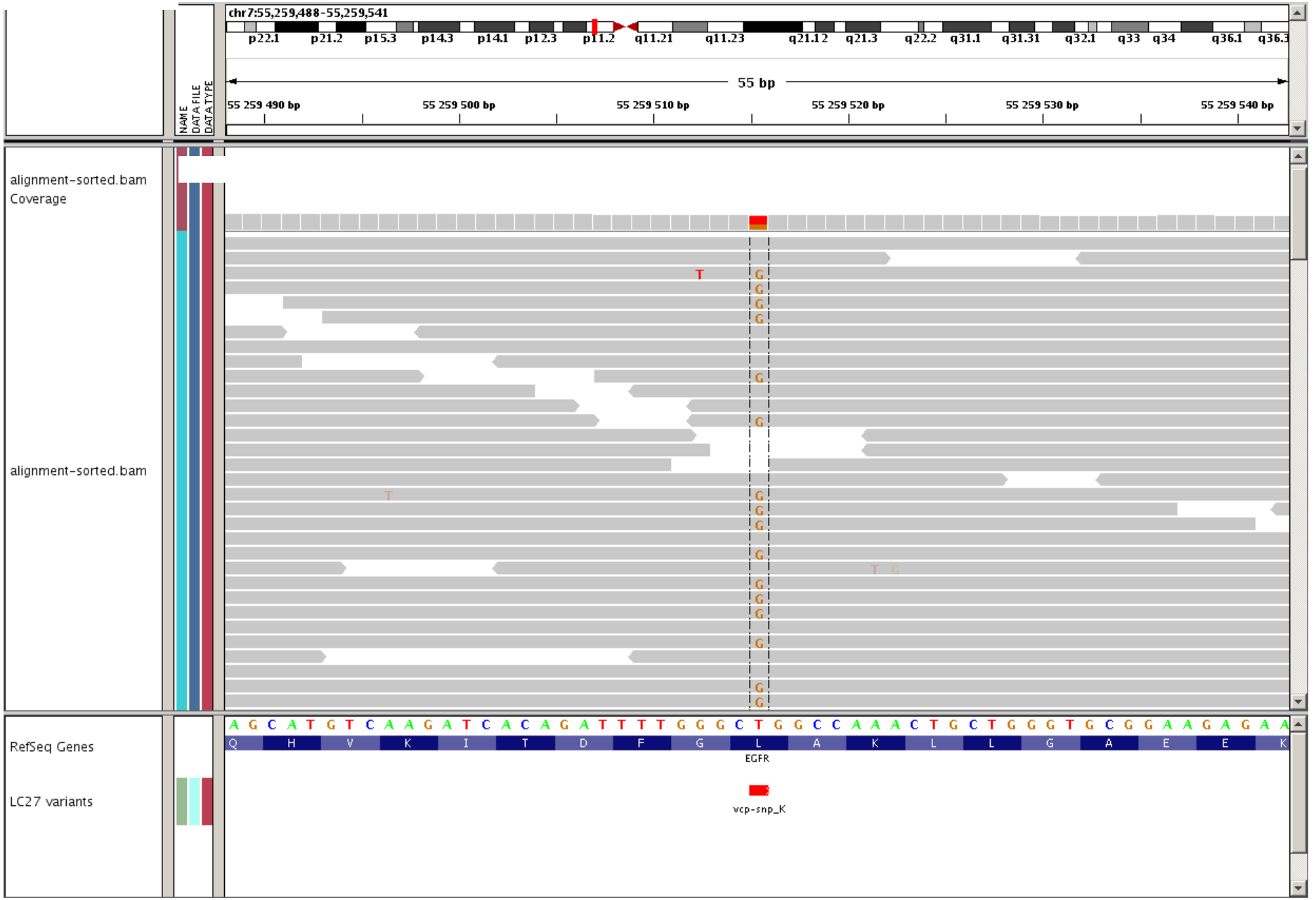
Gene	No of cases compared	Concordant (TDS / PCR) –/–	Concordant (TDS / PCR) +/+	Discordant (TDS / PCR) –/+	Discordant (TDS / PCR) +/–	Percent of concordance	TDS: missense mutations not included in our PCR panel
EGFR							
LAC	81	61	16 (19%)	2[z]	0	97.5	7
CRC	0						
Total	81	61	16	2	0	97.5	7
KRAS							
LAC	78	53	24 (30%)	1[zz]	0	98.7	
CRC	18	16	1	0	1[x]	94.4	
Total	96	69	25	1	1	97.9	
BRAF							
LAC	78	78	0	0	0	100	
CRC	18	16	1	0	0	100	1
Total	96	94	1	0	0	100	1

z One had only 21 reads (poor), the other one 85 reads (OK).
zz 301 reads (very good).
x 94/265 reads, included to PCR-panel (PCR detected 6 out of 7 same mutations)

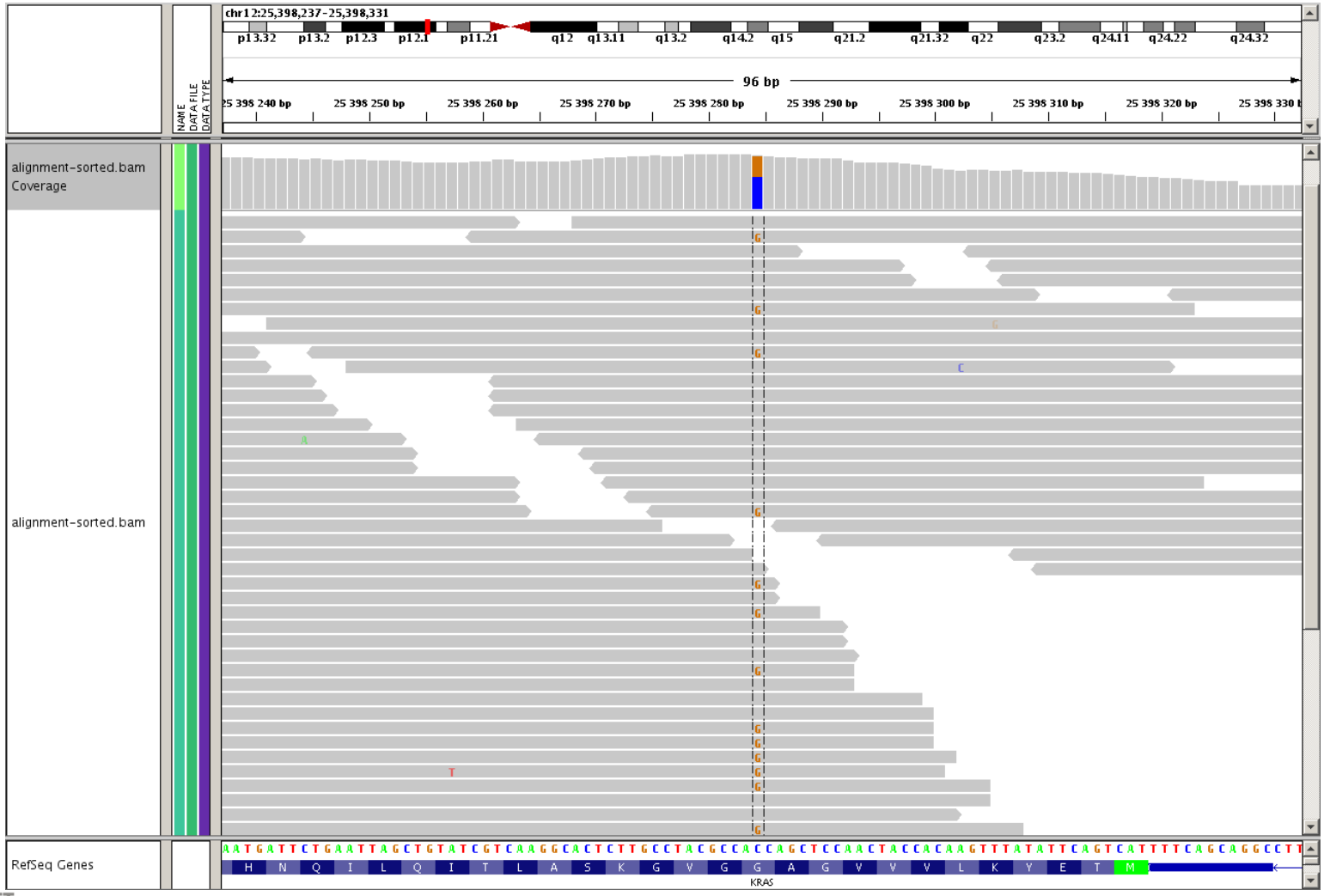
TDS showed in 9 cases a new missense mutation (not included in the PCR test panel)

Gene	Cases	Protein change	Exon	Protein domain
EGFR	1	Val774Met	20	protein kinase
EGFR	1	insertion of SerValGly after codon 767	20	protein kinase
EGFR	1	Leu62Arg	2	extracellular (receptor L domain)
EGFR	1	Ala647Thr	17	transmembrane
EGFR	2	Tyr275Phe Tyr274Phe	7	extracellular (furin-like Cys rich domain)
EGFR	1	rs121913432 * His773Leu	20	protein kinase
EGFR	1	rs148934350 Pro848Leu	21	protein kinase
BRAF *	1	rs121913338 Asp(594)Gly	15	protein kinase

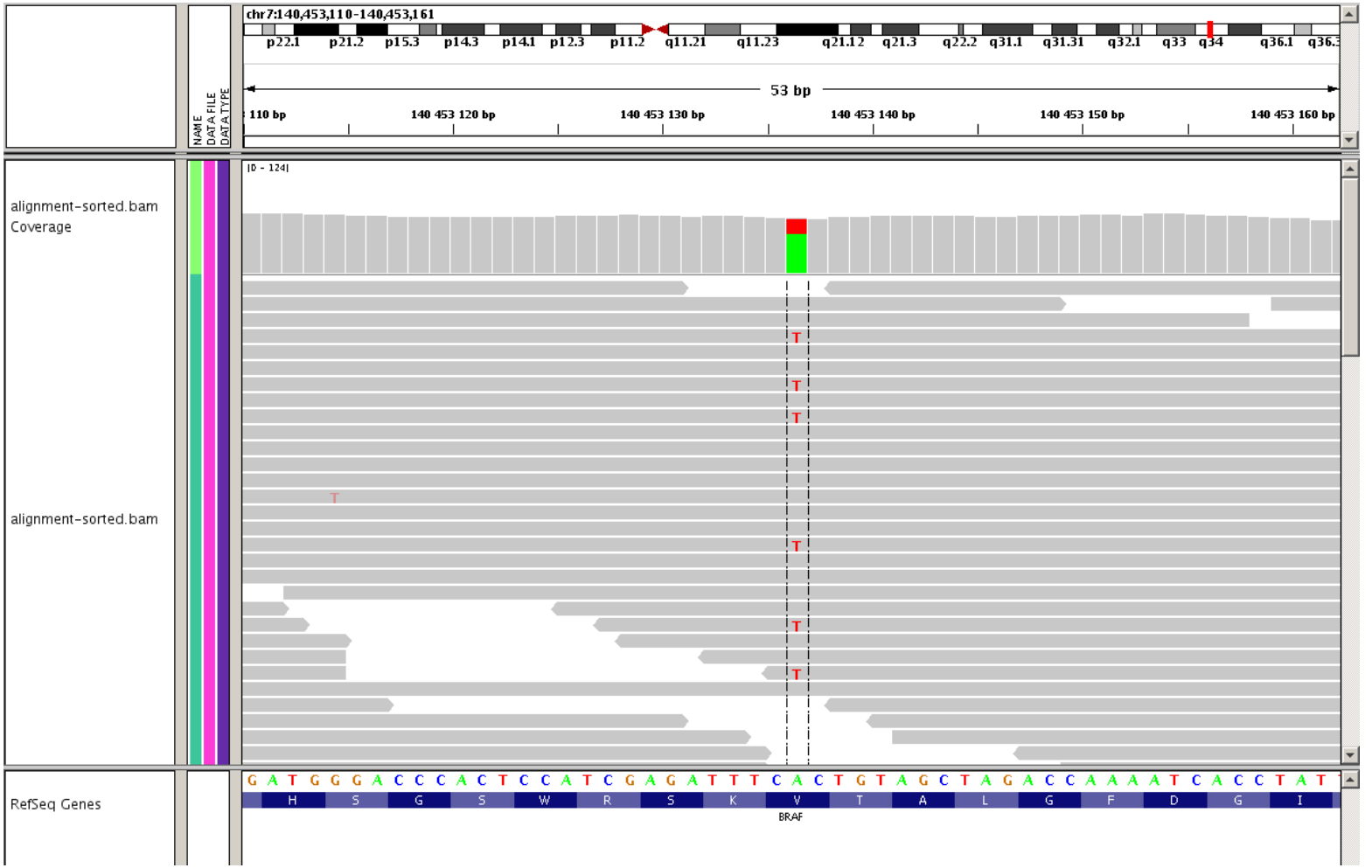
EGFR mutation: Leu858Arg in FFPE LAC sample



FFPE: *KRAS* GLY(12)ALA mutation in LAC



FFPE: *BRAF* mut V600E in CRC



Mutations in 48 TK genes by TDS

Diagnostic validation

EGFR

KRAS

BRAF

ALK/ELM4 -fusion

ABL1, ABL2, AKT1, AKT2, AKT3, ALK, ARAF, AREG, AXL, BAD, BRAF, BTC, CAMK2A, CAMK2B, CAMK2D, CAMK2G, CBL, CBLB, CBLC, CCND1, CDKN1A, CDKN1B, CDKN2A, CRK, CRKL, CSF1R, CTNNB1, DDR1, DDR2, EGF, EGFR, EIF4EBP1, ELK1, EML4, EPHA1, EPHA2, EPHA3, EPHA4, EPHA5, EPHA6, EPHA7, EPHA8, EPHB1, EPHB2, EPHB3, EPHB4, EPHB6, EPHX1, ERBB2 , ERBB3 , ERBB4 , EREG , FGFR1, FGFR2, FGFR3, FGFR4, FLT1, FLT3, FLT4, GAB1 , GRB2, GSK3B, HBEGF, HMGA1, HMGA2, HRAS, IGF1, IGF1R, INSR, INSRR, JUN, KDR, KIT, KRAS, LTK, MAP2K1, MAP2K2, MAP2K4, MAP2K7, MAPK1, MAPK10, MAPK3, MAPK8 , MAPK9, MERTK, MET

MIR125A, MIR126, MIR135B, MIR137, MIR143, MIR152, MIR155, MIR15A, MIR16-1, MIR17HG, MIR181A1, MIR181A2, MIR181D, MIR182, MIR18A, MIR19A, MIR200C , MIR205, MIR206, MIR20A, MIR21, MIR210, MIR221, MIR222, MIR26A1, MIR29A, MIR29C, MIR335, MIR34A, MIR34B, MIR34C, MIR372, MIR378, MIR451, MIR7-1, MIR92A1, MIR98, MIRLET7A1, MIRLET7A2, MIRLET7A3, MIRLET7B, MIRLET7C, MIRLET7D, MIRLET7E, MIRLET7F1, MIRLET7F2, MIRLET7G, MIRLET7I

MST1R, MTOR, MUSK, MYC, NCK1, NCK2, NFE2L2, NRAS, NRG1, NRG2, NRG3, NRG4, PAK1, PAK2, PAK3, PAK4, PAK6, PAK7, PDGFRA, PDGFRB, PDGFRL, PIK3CA, PIK3CB, PIK3CD, PIK3CG, PIK3R1, PIK3R2, PIK3R3, PIK3R5, PLCG1, PLCG2, PRKCA, PRKCB, PRKCG, PTEN, PTK2, PTK7, RAF1, REL, RET, RPS6KB1, RPS6KB2, SHC1, SHC2 , SHC3, SHC4, SOS1, SOS2, SRC, STAT5A, STAT5B, STK11, TGFA , TGFB1, TP53, TRAF6, TYRO3, VEGFA, VEGFB

Red: targeted treatment available (far from complete list)

Recurrent missence mutations in lung carcinoma (more than 20 reads in each). Known missense variations which are clinically significant and probably pathogenic according to the databases

Gene	Cases	Change in protein	Protein domain
INSR	1	Val1012Met Val1000Met	cytoplasmic
KDR *(VEGFR2)*	2	Cys482Arg	Ig-like

Novel missense single nucleotide mutations in LAC

Gene	Cases	Change in protein	Exon	Protein domain
EPHB1	1	Ile55Val	3	ligand binding
EPHB1	1	Val619Ala	10	protein kinase
EPHB6	1	Arg300Pro	7	protein kinase
EPHB6	1	Glu710Lys	15	protein kinase
ERBB2	2	Arg929Gln	23	protein kinase
PDGFRA	1	Lys606Asn	13	protein kinase
PDGFRB	1	Ala848Pro	18	protein kinase

Novel indel mutations in LAC

Gene	Cases	Consequence	Exon and amino acid
ERBB2	1	insertion of 4 codons	ex 20, aa 770
EPHB2	1	2 base deletion; frameshift	ex 14, aa 849
EPHA1	1	deletion of 1 base; frameshift	ex 16, aa 844
ALK	1	deletion of 1 base; frameshift	ex 24, aa 1246
EPHB3	1	deletion of 1 base; frameshift	ex 3, aa 266

Most frequent known single nucleotide variations (more than 14 times more frequent in LC patients compared to the normal human population – 1000 genome project/ES cohort populations)

Gene	Cases	Change in protein	Exon	Normal frequence	Patient frequence
CSF1R	3	Gly747Arg	17	0,0013	0,0183
EPHA3	4	Arg777Gly	13	0,0005	0,0244
ERBB3	5	Leu795Val	20	0,0018	0,0305

What about mutations in genes of non-TKs and miRNAs?

- Numerous known and novel mutations were detected; their confirmations are in progress

ALK fusions

Diagnostic validation

EGFR

KRAS

BRAF

***ALK/ELM4* -fusion**

ABL1, ABL2, AKT1, AKT2, AKT3, ALK, ARAF, AREG, AXL, BAD, BRAF, BTC, CAMK2A, CAMK2B, CAMK2D, CAMK2G, CBL, CBLB, CBLC, CCND1, CDKN1A, CDKN1B, CDKN2A, CRK, CRKL, CSF1R, CTNNB1, DDR1, DDR2, EGF, EGFR, EIF4EBP1, ELK1, EML4, EPHA1, EPHA2, EPHA3, EPHA4, EPHA5, EPHA6, EPHA7, EPHA8, EPHB1, EPHB2, EPHB3, EPHB4, EPHB6, EPHX1, ERBB2 , ERBB3 , ERBB4 , EREG , FGFR1, FGFR2, FGFR3, FGFR4, FLT1, FLT3, FLT4, GAB1 , GRB2, GSK3B, HBEGF, HMGA1, HMGA2, HRAS, IGF1, IGF1R, INSR, INSRR, JUN, KDR, KIT, KRAS, LTK, MAP2K1, MAP2K2, MAP2K4, MAP2K7, MAPK1, MAPK10, MAPK3, MAPK8 , MAPK9, MERTK, MET

MIR125A, MIR126, MIR135B, MIR137, MIR143, MIR152, MIR155, MIR15A, MIR16-1, MIR17HG, MIR181A1, MIR181A2, MIR181D, MIR182, MIR18A, MIR19A, MIR200C , MIR205, MIR206, MIR20A, MIR21, MIR210, MIR221, MIR222, MIR26A1, MIR29A, MIR29C, MIR335, MIR34A, MIR34B, MIR34C, MIR372, MIR378, MIR451, MIR7-1, MIR92A1, MIR98, MIRLET7A1, MIRLET7A2, MIRLET7A3, MIRLET7B, MIRLET7C, MIRLET7D, MIRLET7E, MIRLET7F1, MIRLET7F2, MIRLET7G, MIRLET7I

MST1R, MTOR, MUSK, MYC, NCK1, NCK2, NFE2L2, NRAS, NRG1, NRG2, NRG3, NRG4, PAK1, PAK2, PAK3, PAK4, PAK6, PAK7, PDGFRA, PDGFRB, PDGFRL, PIK3CA, PIK3CB, PIK3CD, PIK3CG, PIK3R1, PIK3R2, PIK3R3, PIK3R5, PLCG1, PLCG2, PRKCA, PRKCB, PRKCG, PTEN, PTK2, PTK7, RAF1, REL, RET, RPS6KB1, RPS6KB2, SHC1, SHC2 , SHC3, SHC4, SOS1, SOS2, SRC, STAT5A, STAT5B, STK11, TGFA , TGFB1, TP53, TRAF6, TYRO3, VEGFA, VEGFB

Red: targeted treratment available (far from comlete list)

Concordance of the results of ALK fusions by TDS, FISH, PCR, and IHC in LAC

- Results from 95 formalin-fixed paraffin-embedded tumor tissue specimens from 87 patients with non-small cell lung carcinoma showed 100 percent concordance: six positive and 81 negative specimens were detected.

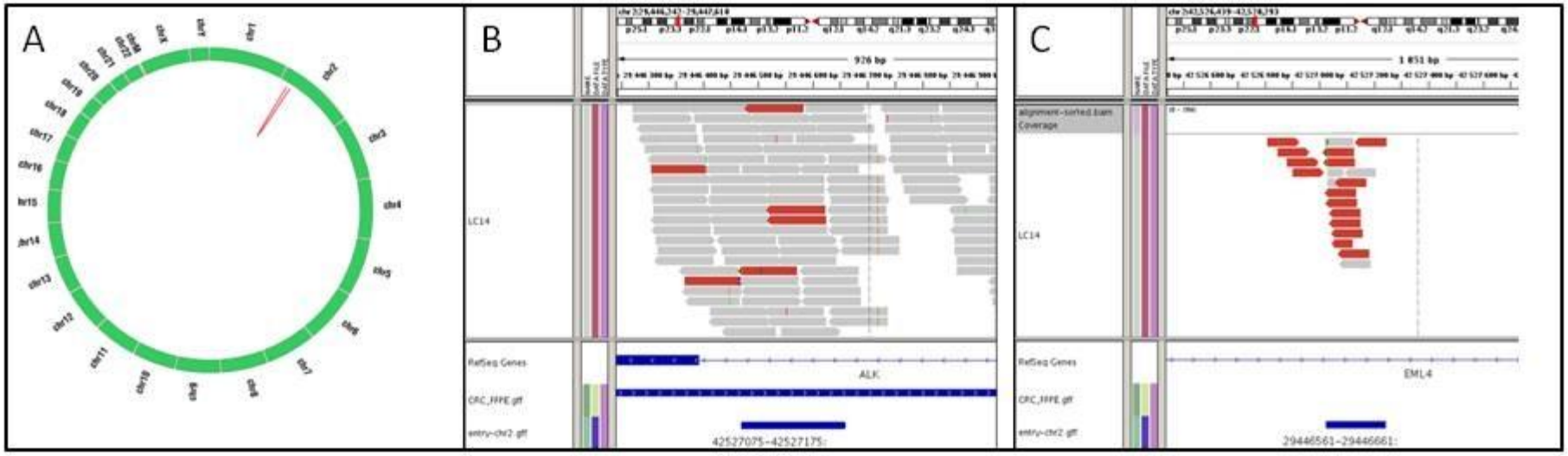
ALK fusion confirmation

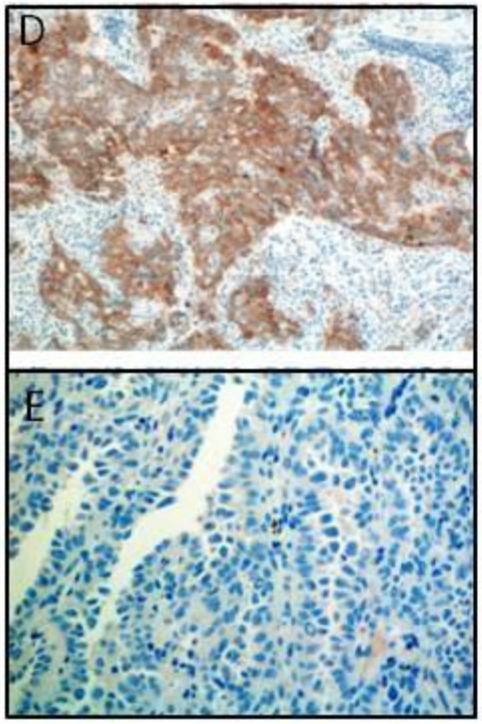
- FISH: Vysis ALK Break Apart FISH probe kit
- PCR: AmoyDx EML4-ALK Fusion Gene Detection Kit
- IHC: Novocastra clone 5A4 by BenchMark XT

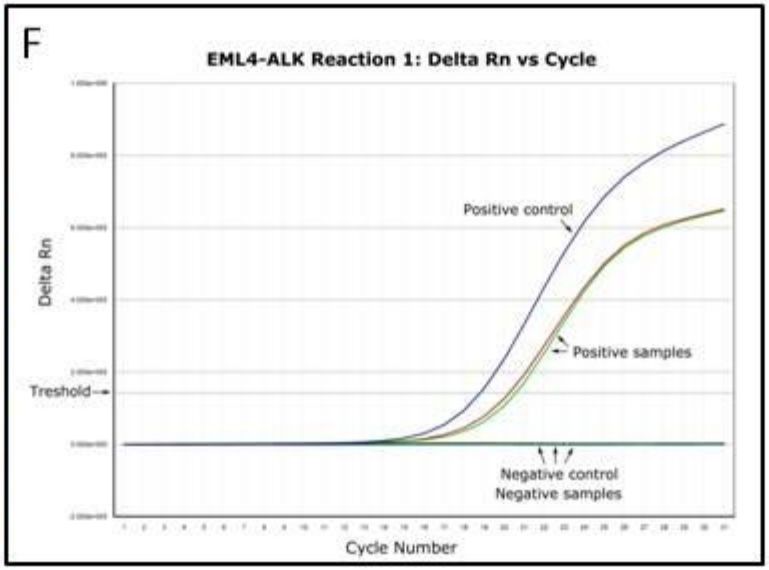
Positive cases detected by FISH, IHC, RT-PCR, and targeted next-generation sequencing (NGS)

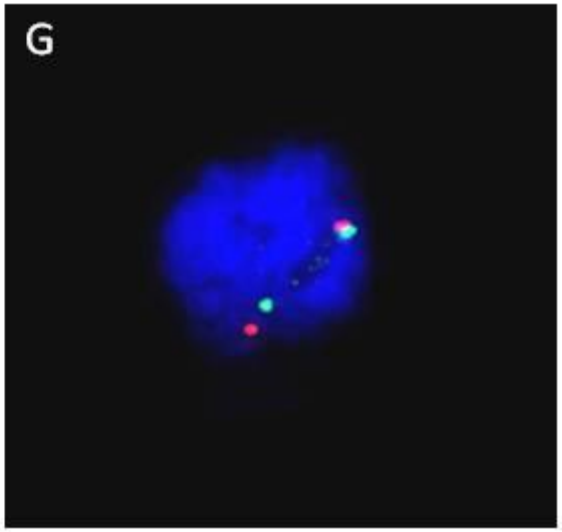
Sample ID	Age	Sex	Smoking	FISH	IHC	RT-PCR	NGS
				Fusion type / % of positive cells		Variant type	Fusion /break position in ALK
LC1	46	F	Non-smoker	Deletion /70%	Strongly positive	*EML4-ALK* variants 1/2/3a/3b	chr2: 29447508–29447608
LC11	48	M	Non-smoker	Inversion /46%	Strongly positive	*EML4-ALK* variants 1/2/3a/3b	chr2:29446498–29446566
LC14*	44	M	Non-smoker	Inversion/74%	Not studied	*EML4-ALK* variants 1/2/3a/3b	chr2:29446561-29446661
LC74*	44	M	Non-smoker	Inversion/46%	Strongly positive	*EML4-ALK* variants 1/2/3a/3b	chr2:29446561-29446661
LC51	51	M	Non-smoker	Deletion/66%	Strongly positive	*EML4-ALK* variants 1/2/3a/3b	chr2:29446527–29446625
15915	63	M	Non-smoker	Deletion /40%	Strongly positive	*EML4-ALK* variants 1/2/3a/3b	Not studied

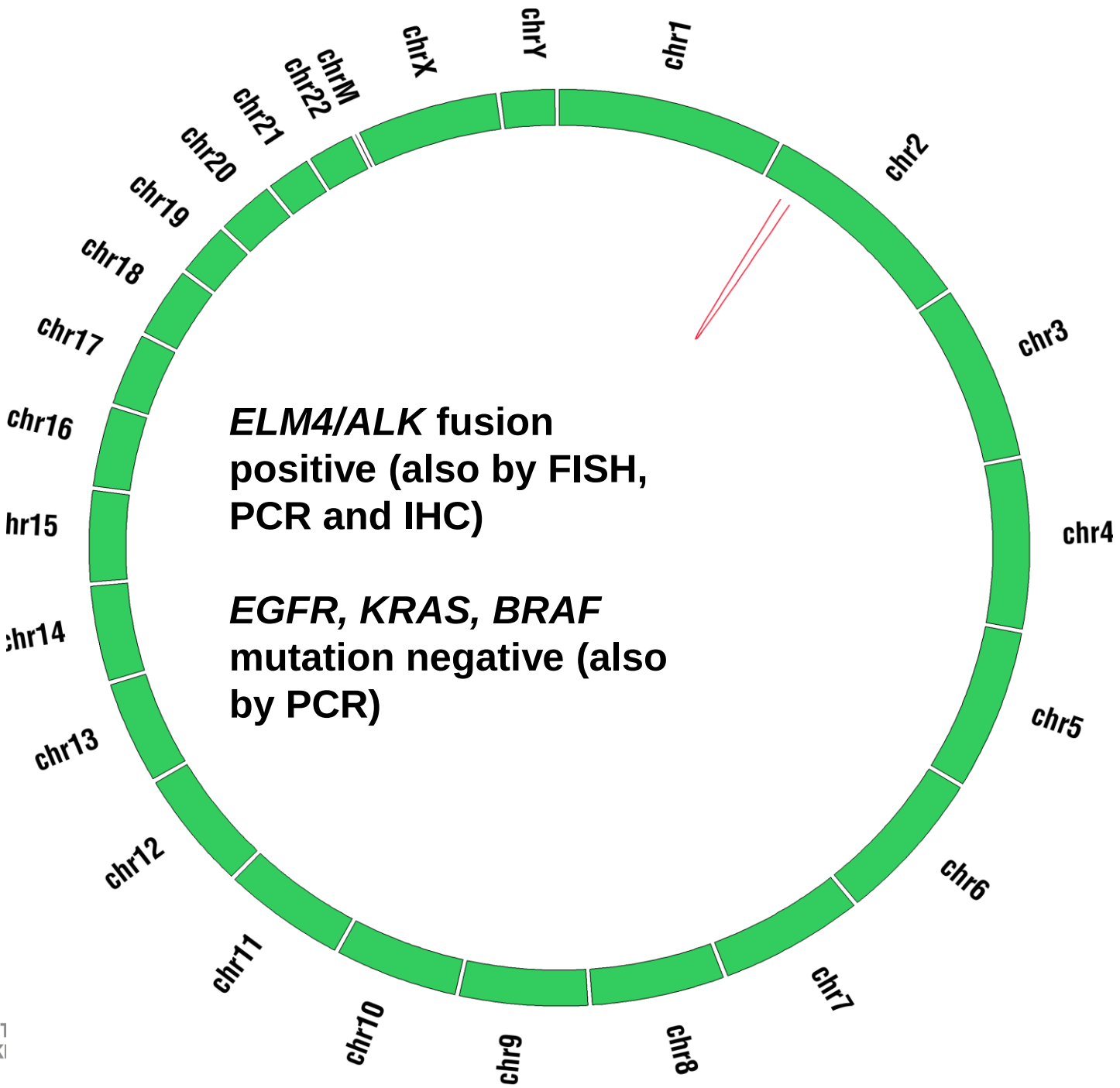
*LC14 and LC74 are samples from the same patient but from a different FFPE block.


A
B
C
926 bp
1 851 bp
LC14
RefSeq Genes
ALK
EML4
CRC_FFPE.gff
entry-chr2.gff
42527075-42527175:
29446561-29446661:


D
E


F
EML4-ALK Reaction 1: Delta Rn vs Cycle
Delta Rn
Cycle Number
Positive control
Positive samples
Treshold
Negative control
Negative samples

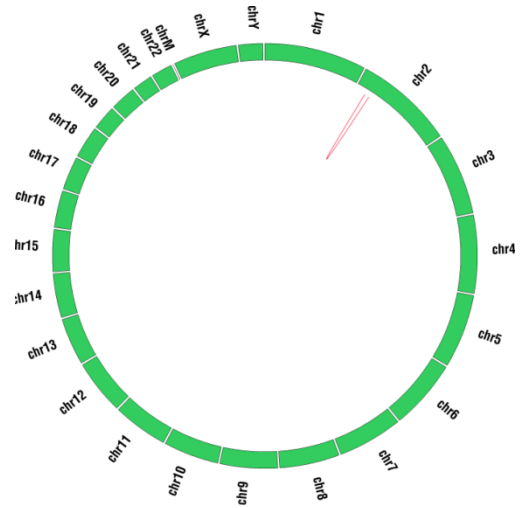

G

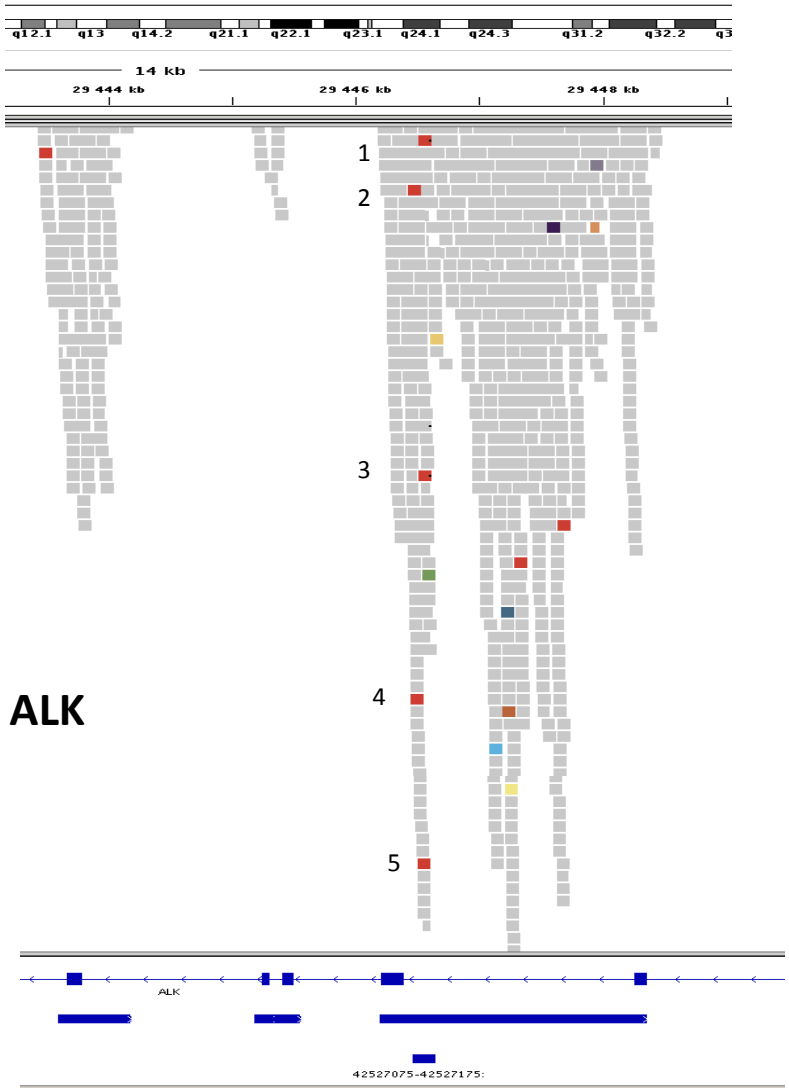


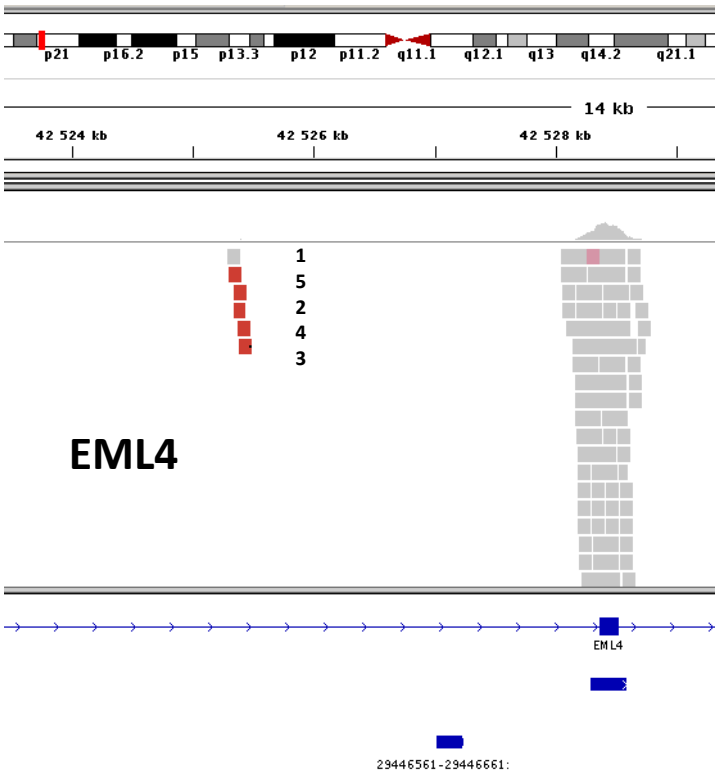
YLIOPISTO
RS UNIVERSIT
' OF HELSINKI

LC51

- ***ELM/ALK*** fusion positive (also by FISH, 66%, PCR and IHC)
- *EGFR, KRAS* and *BRAF* mutation negative (also by PCR)



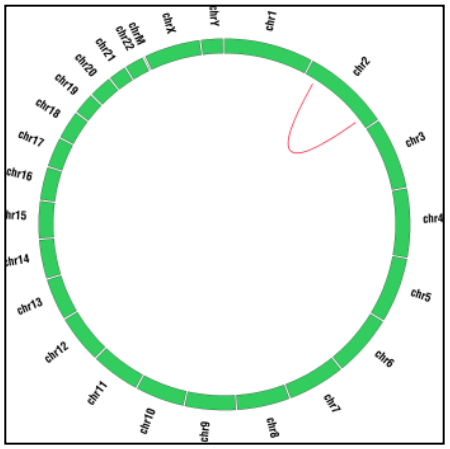


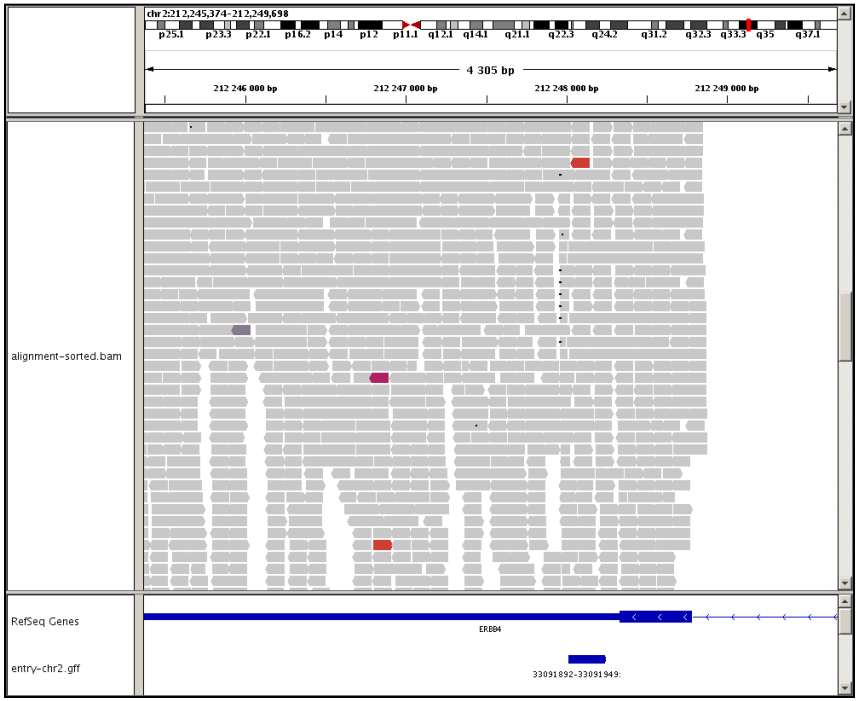


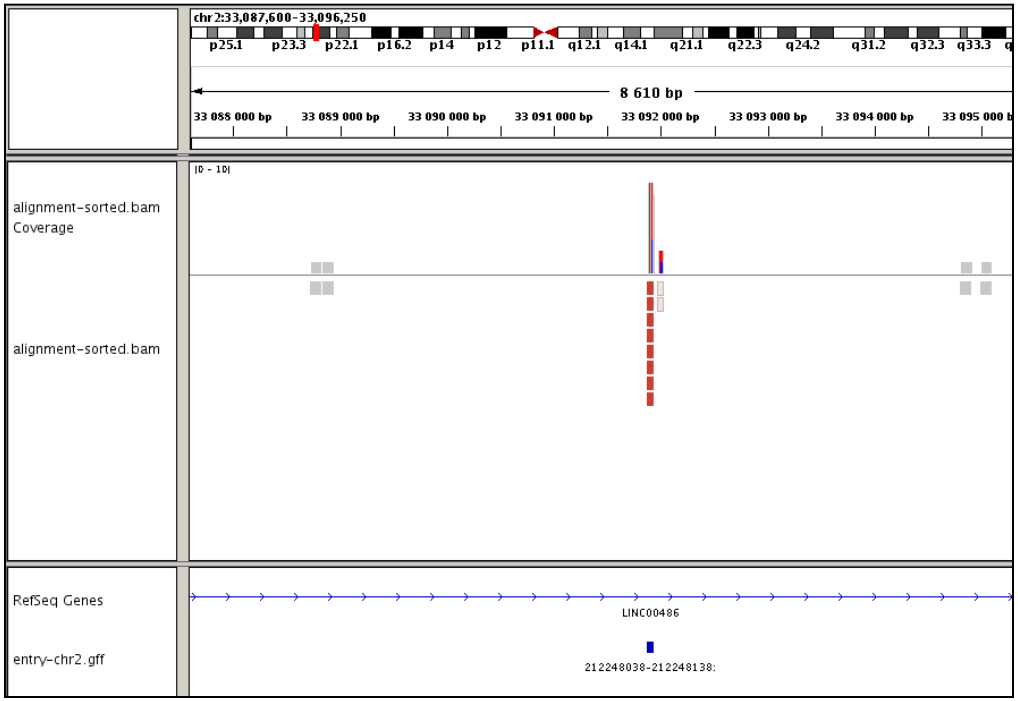
Novel ALK or other fusions?

- No novel ALK
- Other fusion
  - Some novel fusions described; their confirmation by other methods are in progress

LC-27: *ERBB4/LINC00486* fusion at chromosome 2



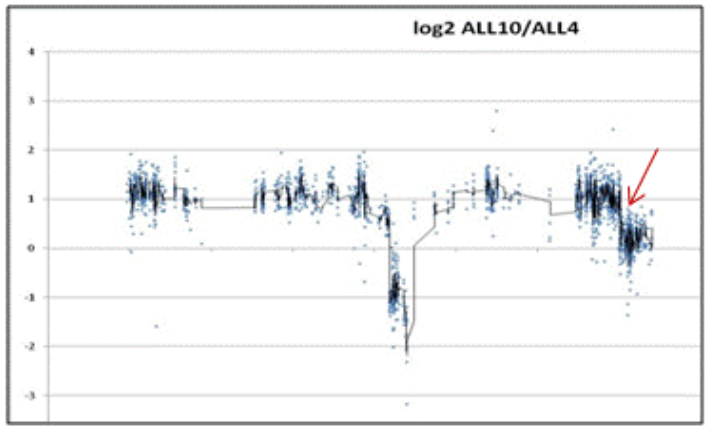


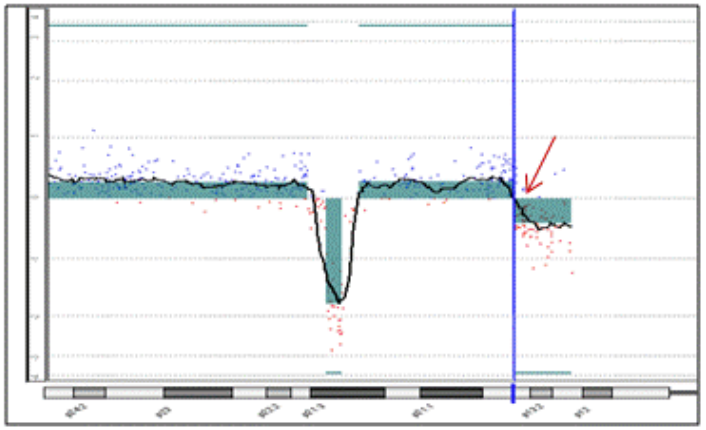


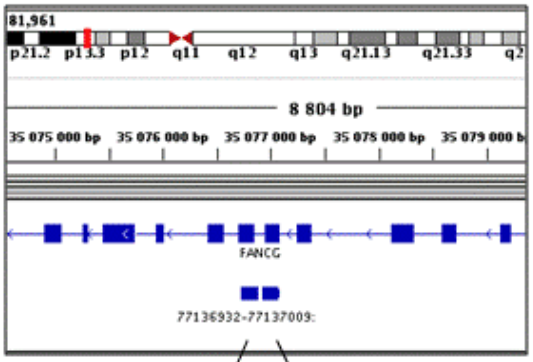
Advantages and drawbacks of TDS

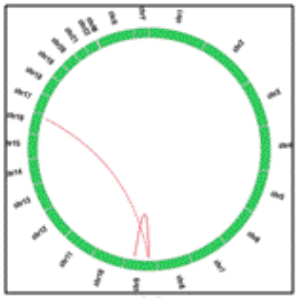
Comparison of the methods used for mutation and *ALK* fusion detection in present study

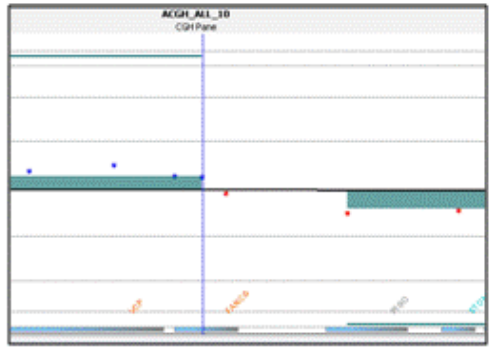
	FISH (ALK fusion)	**IHC (ALK fusion)**	**RT-PCR (all)**	**NGS (all)**
Specificity	No fusion specification	No fusion specification	Only targeted mutations and ALK-ELM4 fusions	All kind of mutations and fusions
Sensitivity	5-10%	5-10%	1-5%	5-10%?
Time used for analysis	2-3 days	Half a day	1 day	10 days
Cost	Medium (~200 euros)	**Low** (~20 euros)	Medium (~150 euros)	High (~1000 euros)
Is FFPE material applicable?	Yes	Yes	Not always	Yes
Amount of material required	**One tissue section** (2.5µm thick)	One tissue section (2.5µm thick)	0.1-5 µg of RNA	2-3 µg of DNA
Possibility to see large range of gene mutations in one analysis	No	No	No	**Yes**
Possibility to detect mutations, fusions and copy number changes simultaneously	No	No	No	**Yes**
The applicability to average laboratory of pathology	For most of the laboratories	For all laboratories	For some of the laboratories	Only for some laboratories

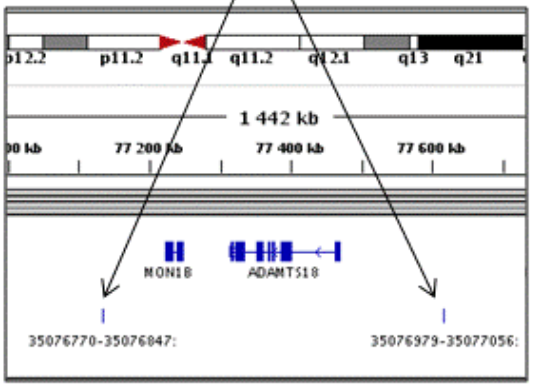

A
log2 ALL10/ALL4

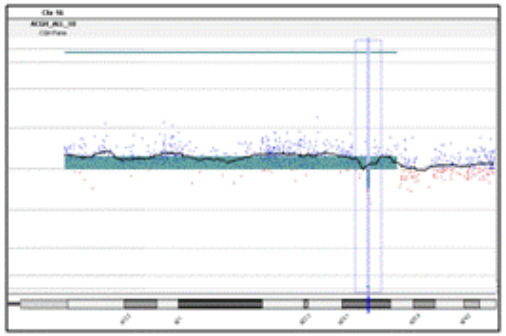

B

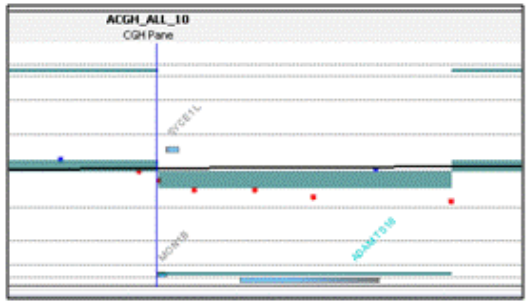

C
81,961
p21.2 p13.3 p12 q11 q12 q13 q21.13 q21.33
8 804 bp
35 075 000 bp 35 076 000 bp 35 077 000 bp 35 078 000 bp 35 079 000 b
FANCG
77136932-77137009:


H


E
ACGH_ALL_10
CGH Pane


D
p12.2 p11.2 q11.1 q11.2 q12.1 q13 q21
1 442 kb
77 200 kb 77 400 kb 77 600 kb
MON1B
ADAMTS18
35076770-35076847:
35076979-35077056:


F


G
ACGH_ALL_10
CGH Pane

Thus, why TDS?

- Single test - no need for multiple testing
- Screen all mutations in all relevant genes
- Saves time
- Saves material

Thus, why TDS (cont.)

- New activating mutations in the gene are found out
- New resistant mutations are found out
- As a whole, mechanisms for drug resistance and molecular pathogenesis of cancer are learnt

Projects in progress

- Validation of NGS (exom sequencing some new kits, too) for diagnostic practice in cancers
- Clinically relevant mutations (biomarkers) by NGS in lung carcinomas and mesothelioma
- Asbestos associated mutations (biomarkers) by NGS in lung carcinomas and mesothelioma

Thank you!

- For support: Finnish Academy, Jusélius Foundation, Finnish Cancer Organizations, EVO, Pfizer, Amgen, Astra Zeneca, Lab21, Abbott Molecular Inc.
- For comfortable cooperation:

  FISH: Milja Tikkanen, Emmi Hokkanen, Tiina Wirtanen

  PCR: Katja Merkkiniemi, Soili Kytölä, Aino Teleranta Keerie, Stuart Bloor

  NGS: Virinder Sarhadi, Aino Wirtanen, Pekka Ellonen, Satu Mäki-Nevala

  IHC: Satu Remes

  Pathology: Kaisa Salmenkivi, Mikko Rönty

  Oncology:: Aija Knuuttila, Milla Katajisto, Pia Österlund

  Bioinformatics: Leo Lahti, Ilari Scheinin