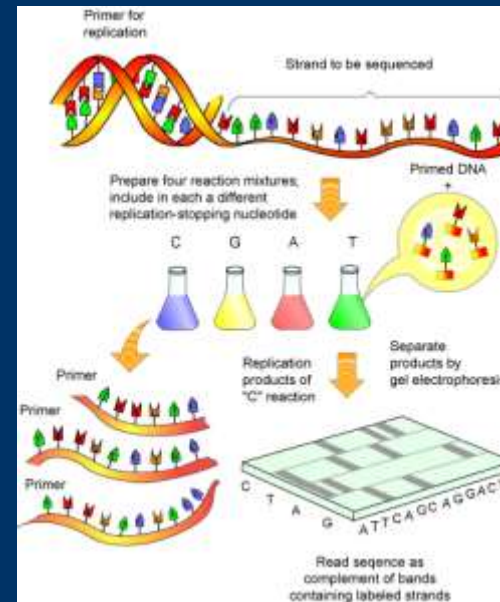
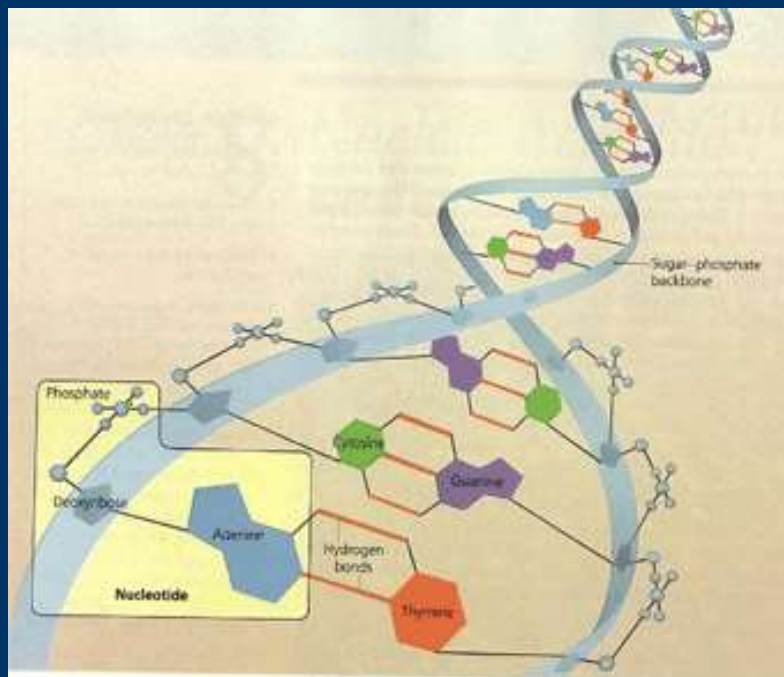


Μέθοδοι και εφαρμογές Διερεύνησης Νουκλεοτιδικής Αλληλουχίας (Sequencing)



Α. Σαέττα Καθηγήτρια
Α. Εργ. Παθολογικής Ανατομικής, ΕΚΠΑ

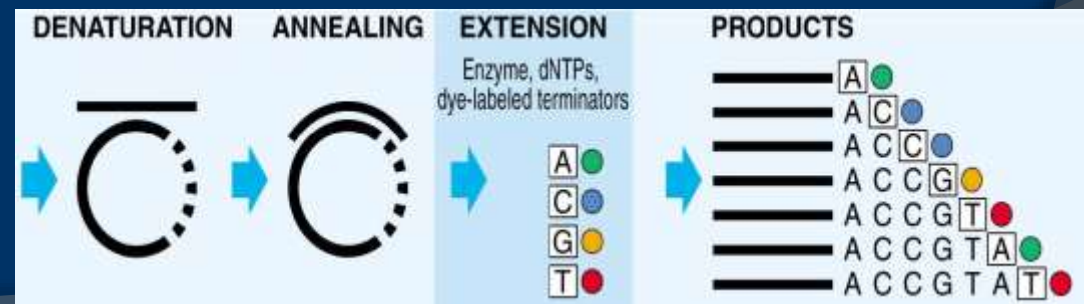
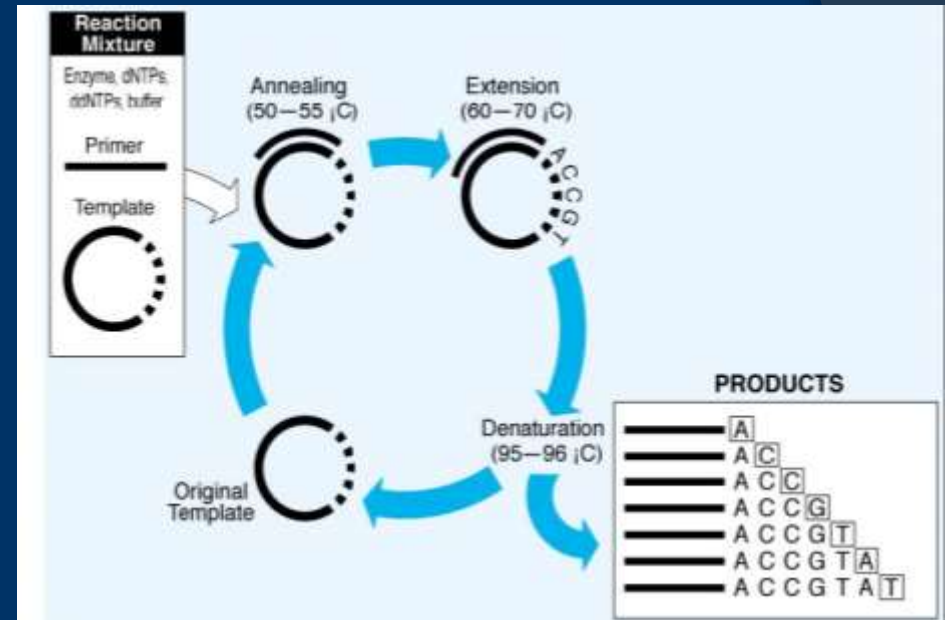
Αλληλούχιση

- DNA sequencing- βασικές μέθοδοι και τεχνολογίες της μοριακής που οδηγούν στο διάβασμα της αλληλουχίας βάσεων σε ένα μόριο DNA (αδενίνη, γουανίνη, κυτοσίνη, θυμίνη) και άρα την ανάλυση γονιδίων σε νουκλεοτιδικό επίπεδο. Έχει πληθώρα εφαρμογών στην έρευνα , διαγνωστική, βιοτεχνολογία .
- Ορισμός. Ο προσδιορισμός της ακριβούς αλληλουχίας νουκλεοτιδίων σε ένα δείγμα DNA. Στάδιο απαραίτητο για ανίχνευση μεταλλαγών
- Αρχικό υλικό : προϊόντα PCR ή κλωνοποιημένα γονίδια

Μέθοδος αλληλούχισης κατά Sanger ή Dideoxy ή τερματισμού της αλυσίδας

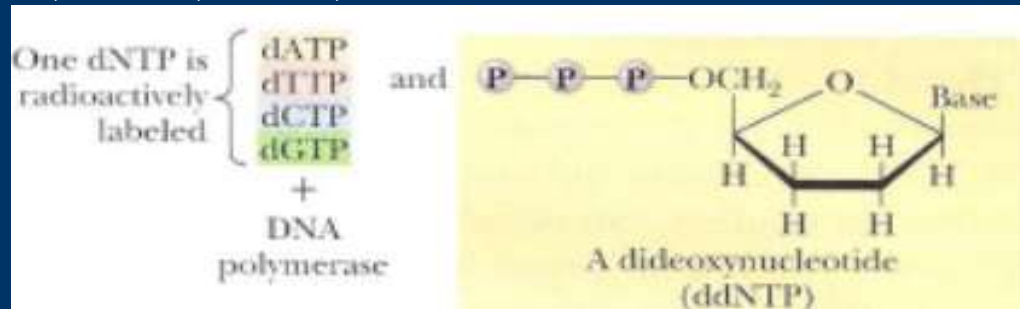
● Στάδια:

1. Αποδιάταξη
2. Πρόσδεση εκκινητή αλληλούχισης
3. Επιμήκυνση της αλυσίδας
4. Ηλεκτροφόρηση



Μέθοδος αλληλούχισης κατά Sanger Chain-termination Method

Η αντίδραση επιμήκυνσης καταλύεται από το ένζυμο DNA πολυμεράση και απαιτούνται τα τέσσερα τριφωσφορικά δεοξυριβονουκλεοτίδια dNTPs: dATP, dCTP, dTTP, dGTP



Επίσης προστίθενται τροποποιημένα νουκλεοτίδια που ονομάζονται δι-δεοξυνουκλεοτίδια ddNTPS.

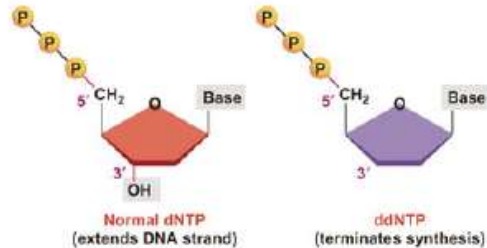
Αυτά ενσωματώνονται κανονικά στην νεοσυντιθέμενη αλυσίδα αλλά εμποδίζουν την περαιτέρω επιμήκυνση της. Λόγω της χημικής δομής τους δεν έχουν -OH στο 3 της ριβόζης έτσι ώστε το επόμενο νουκλεοτίδιο δεν μπορεί να προσδεθεί και να σχηματιστεί ο φωσφοδιεστερικός δεσμός.

Η επιμήκυνση της αλυσίδας σταματά κατά την ενσωμάτωσή τους

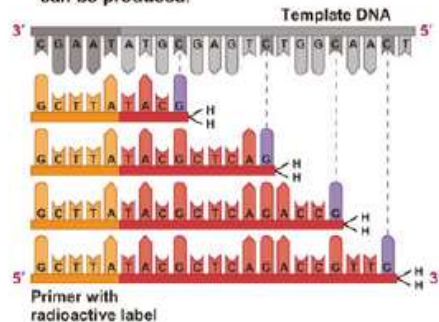
Add enough so each ddNTP is randomly and completely incorporated at each base

Classical Sanger dideoxy sequencing method

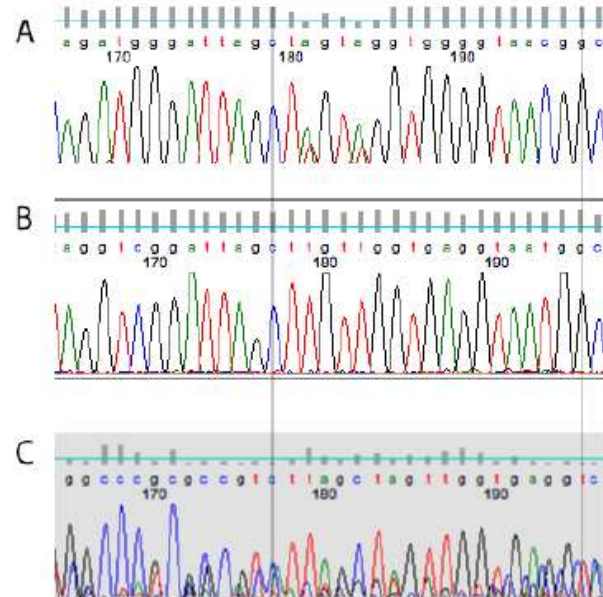
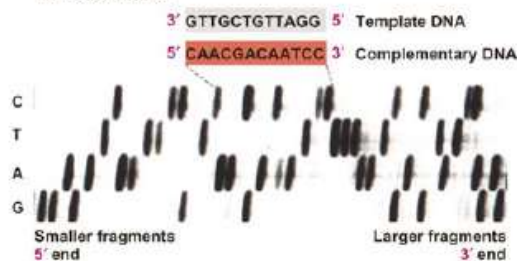
(a) ddNTPs terminate DNA synthesis.



(b) Using ddNTPs, daughter strands of different length can be produced.



(c) Different-length strands can be lined up by size to determine DNA sequence.



Dideoxynucleotide sequencing represents only one method of sequencing DNA. It is commonly called Sanger sequencing since Sanger devised the method. This technique utilizes 2',3'-dideoxynucleotide triphosphates (ddNTPs), molecules that differ from deoxynucleotides by the having a hydrogen atom attached to the 3' carbon rather than an OH group. These molecules terminate DNA chain elongation because they cannot form a phosphodiester bond with the next deoxynucleotide.

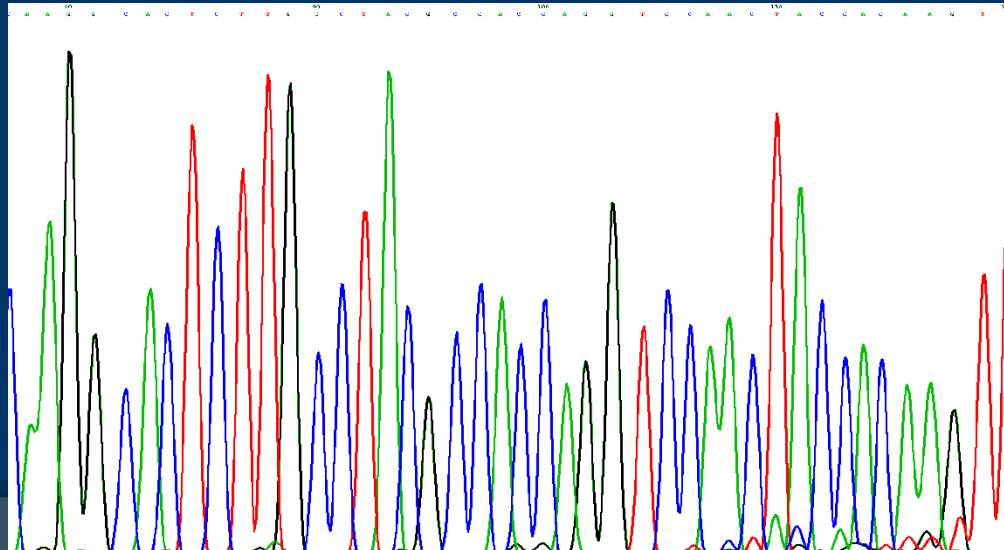
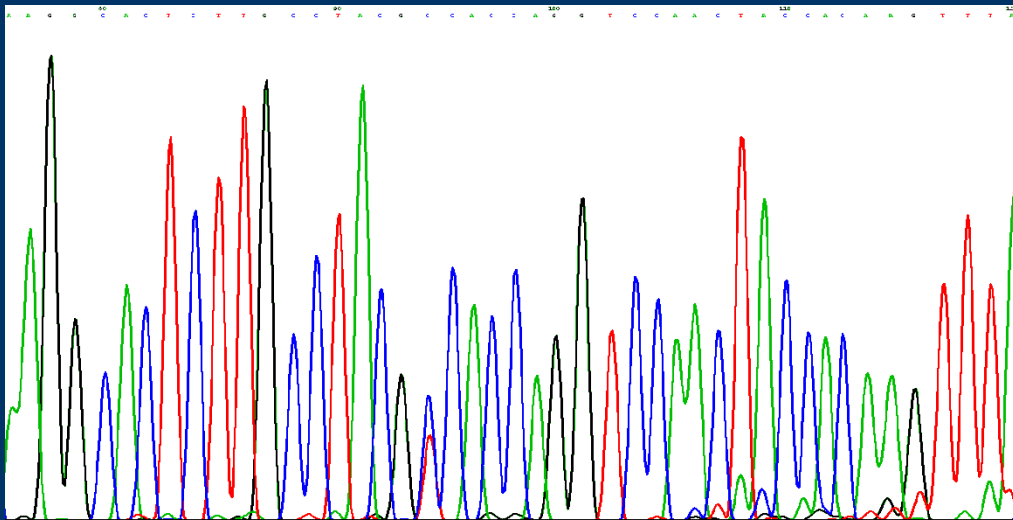
Αλληλούχιση κατά Sanger

Περιορισμοί της μεθόδου

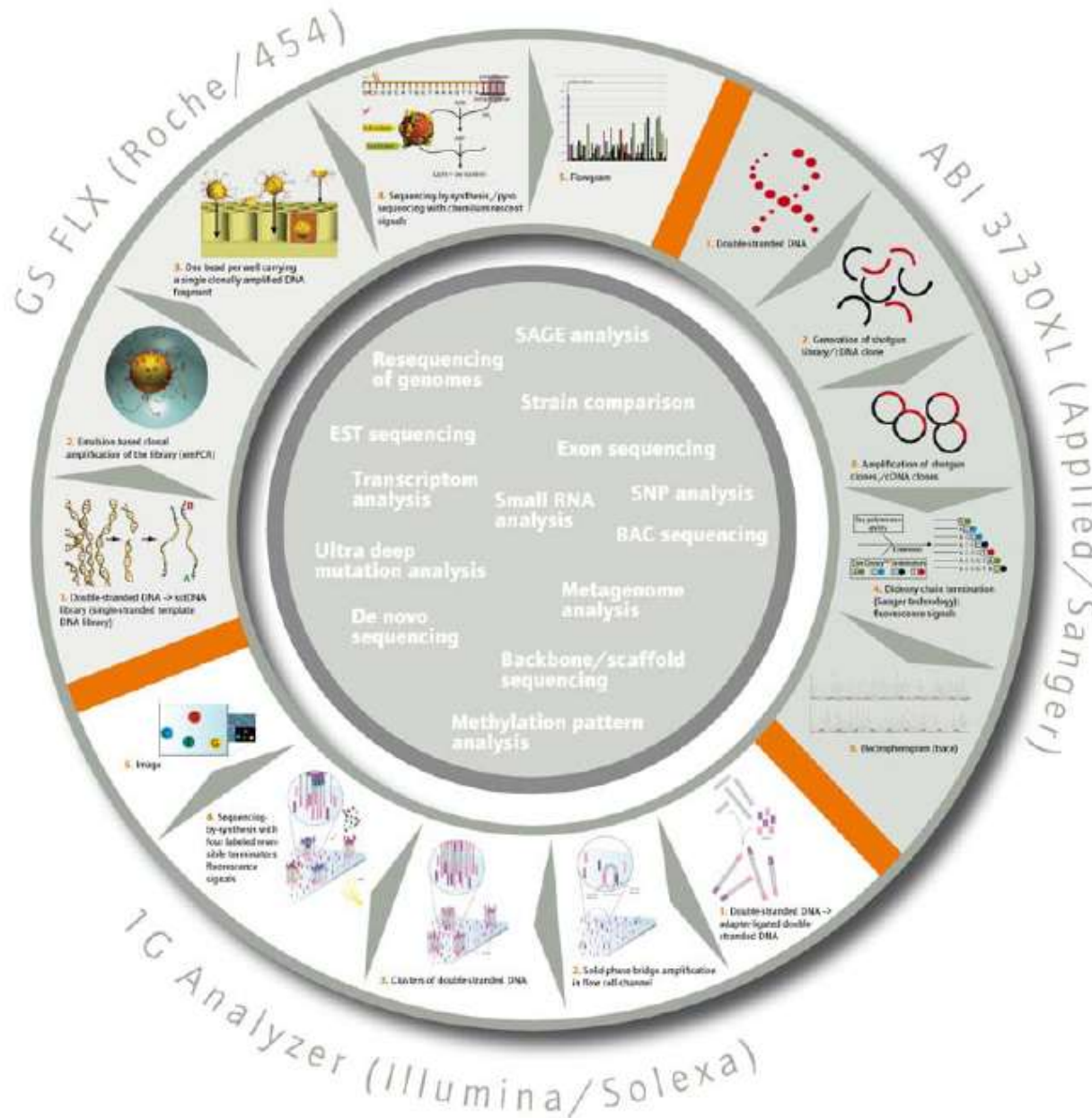
- Διαβάζει 500-750bp (το ανθρώπινο γονιδίωμα υπολογίζεται 3 δισ bp)
- Χρονοβόρος διαδικασία

Αλληλούχιση κατά Sanger

Εφαρμογές: Ανίχνευση μεταλλαγών, πολυμορφισμών



Next generation sequencing



Sanger Institute

- 1992-1999 Sanger seq.

ABI 373/377 slab gel sequencer
2-3runs/day 36-96 samples run
100kb/ instrument /day
80 staff to operate

- 2000

ABI 3700 capillary Sequencer
8runs/day 96 samples/run
400kb /instrument /day
40staff to operate

- 2004

ABI 3730XL Capillary Sequencer
15-40runs/day 96samples/run
1-2MB/instrument/day
4staff to operate

2003 2,76Gb/month

Aug 2008- 278Gb

Aug 2009 1,38Tb

2010 Illumina 200GB/run

60x depth /run

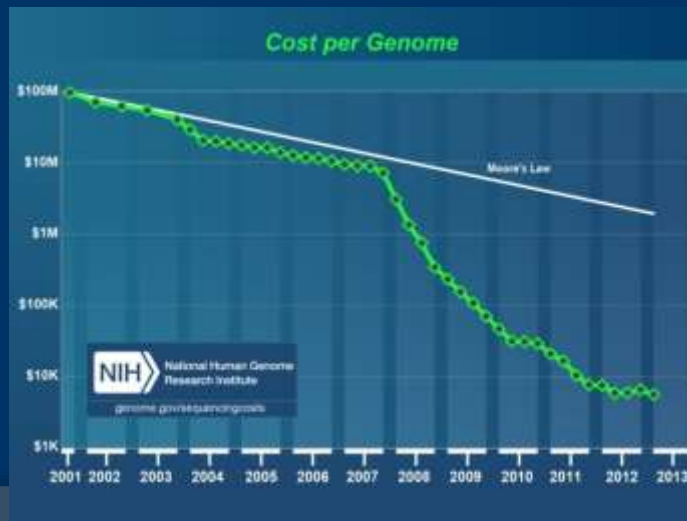
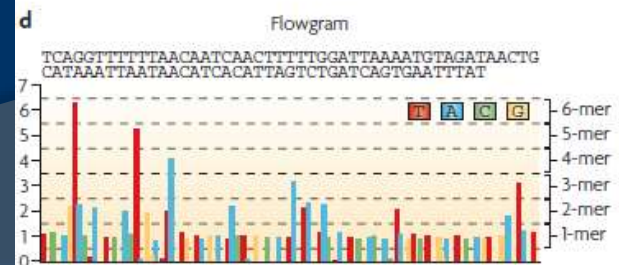
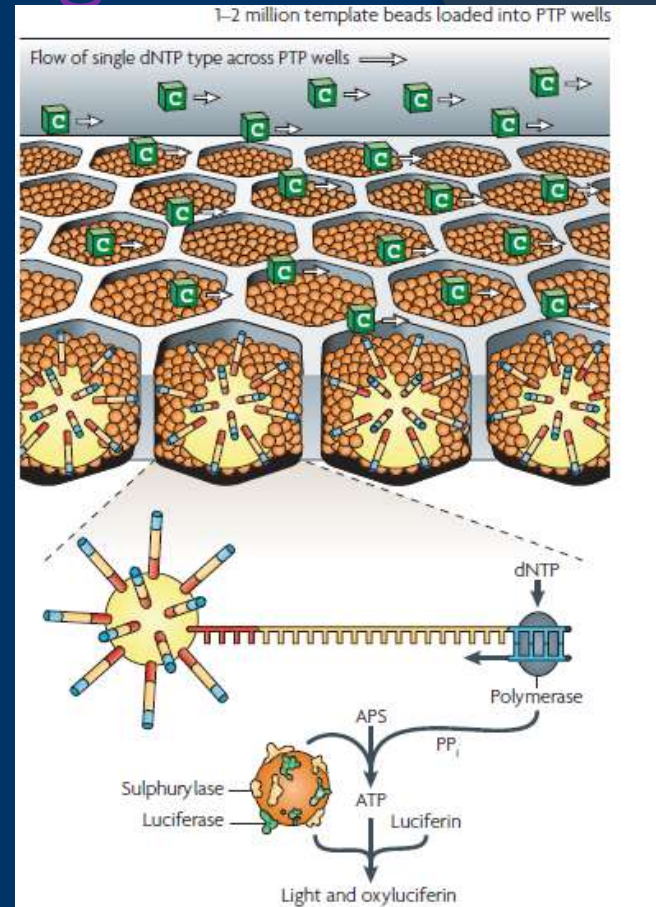


Next generation DNA sequencing - NGS

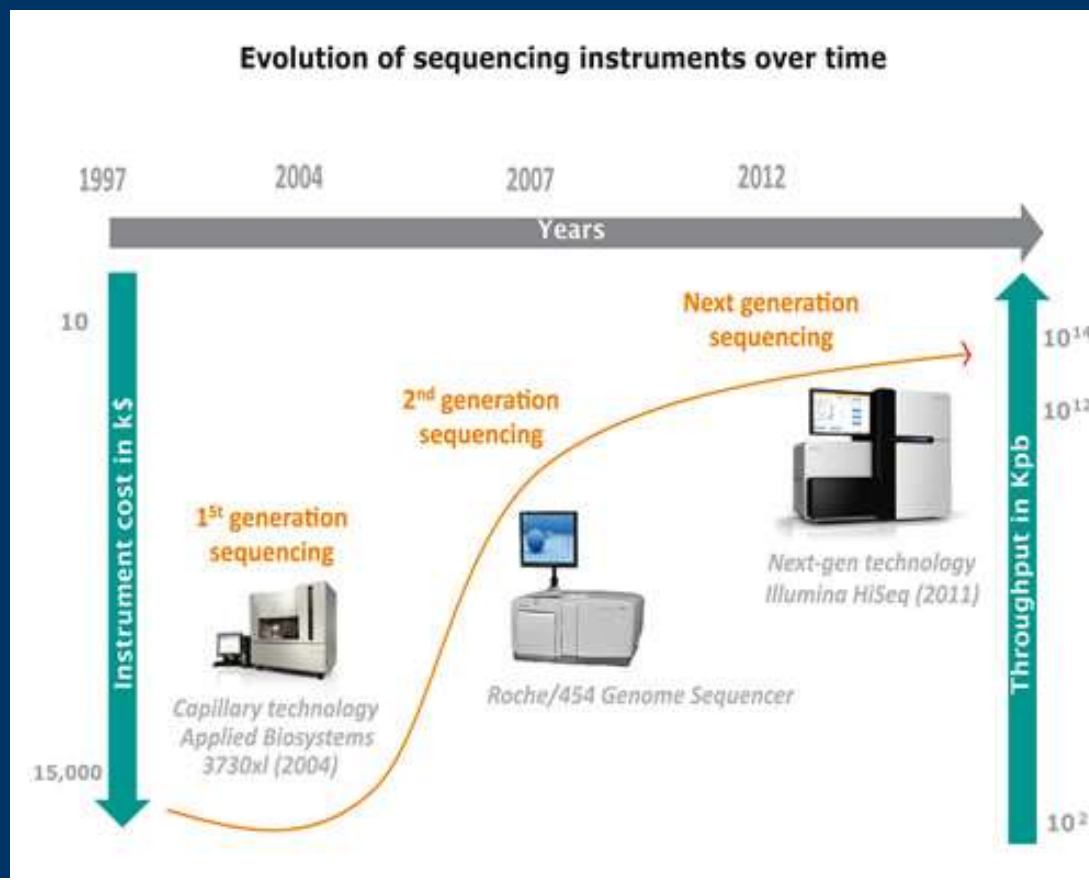
- Μαζικό παράλληλο διάβασμα αλληλουχιών DNA (Massively parallel DNA sequencing)
- Πολύπλοκες τεχνολογίες → χρησιμοποιούν περίπλοκη χημεία → ενζυμολογία, οπτικά υψηλής ανάλυσης εξοπλισμό και λογισμικό
- Ανάλυση μονόκλωνων θραυσμάτων από μία βιβλιοθήκη θραυσμάτων
- βιβλιοθήκη= δημιουργείται με προσθήκη κατάλληλων συνδετικών αλληλουχιών προσαρμοστών (linker-adapters) ειδικών για κάθε πλατφόρμα σε θραύσματα που προέρχονται από το DNA που θέλουμε να αναλύσουμε
- Η παρουσία αυτών των linkers δίνει τη δυνατότητα ειδικής ενίσχυσης των θραυσμάτων με PCR
- Illumina
- Thermo/ Ion Torrent
- Pacific Biosciences
- Nanopore

Next generation DNA sequencing - NGS

- Massively parallel DNA sequencing
Millions of simultaneous reactions producing one read from one input molecule
- Technology has advanced dramatically in recent years, allowing for the human genome to be sequenced rapidly at the cost of less than \$5,000, outpacing Moore's Law



NGS MULTIPLE PLATFORMS



Illumina HiSeq 2500



Ion Torrent Proton



Illumina MiSeq



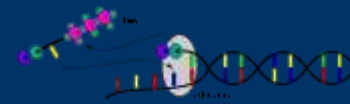
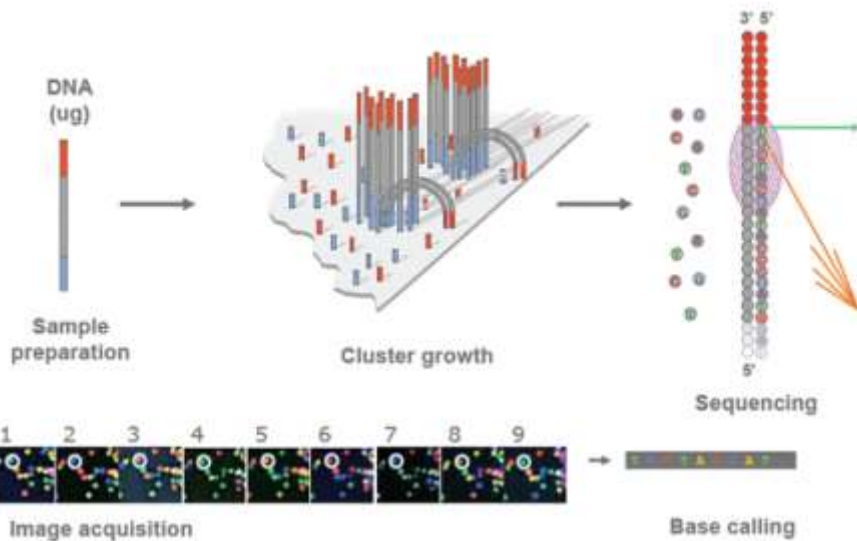
Ion Torrent PGM



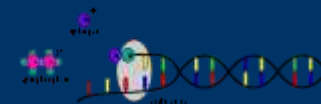
NGS DIFFERENT CHEMISTRIES

Ion semiconductor chemistry/
Ion Torrent

Illumina Sequencing Technology



Polymerase integrates a nucleotide.



Hydrogen and pyrophosphate are released.



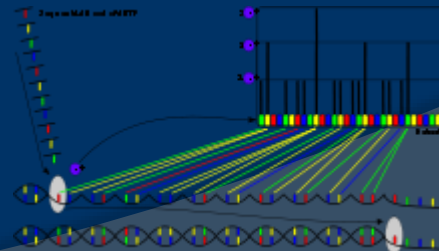
The nucleotide is integrated into the DNA strand, and hydrogen and pyrophosphate are released.



The nucleotide is integrated into the DNA strand, and hydrogen and pyrophosphate are released.

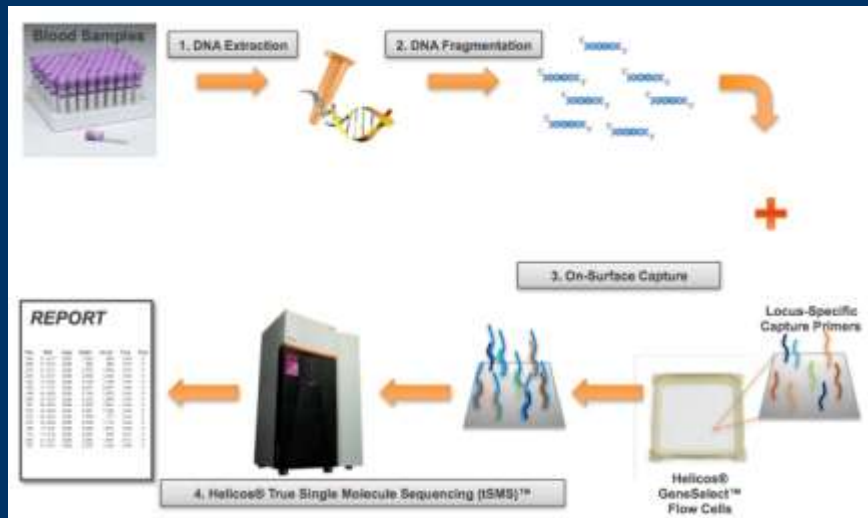


The nucleotide is integrated into the DNA strand, and hydrogen and pyrophosphate are released.

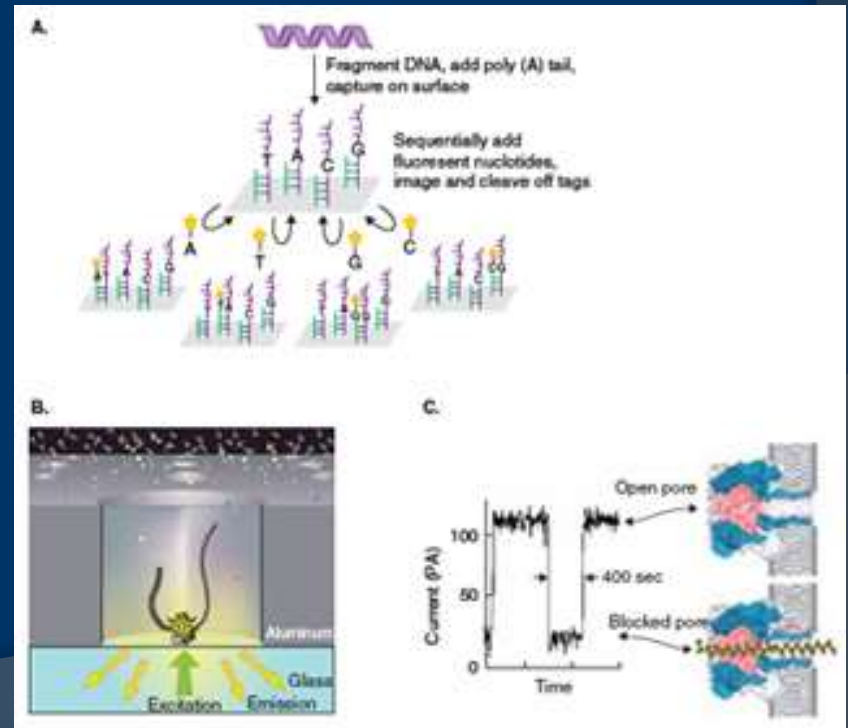


NGS DIFFERENT CHEMISTRIES

Helicos Single Molecule Sequencing



Pacific Biosciences Single Molecule Seq



454/FLX ROCHE

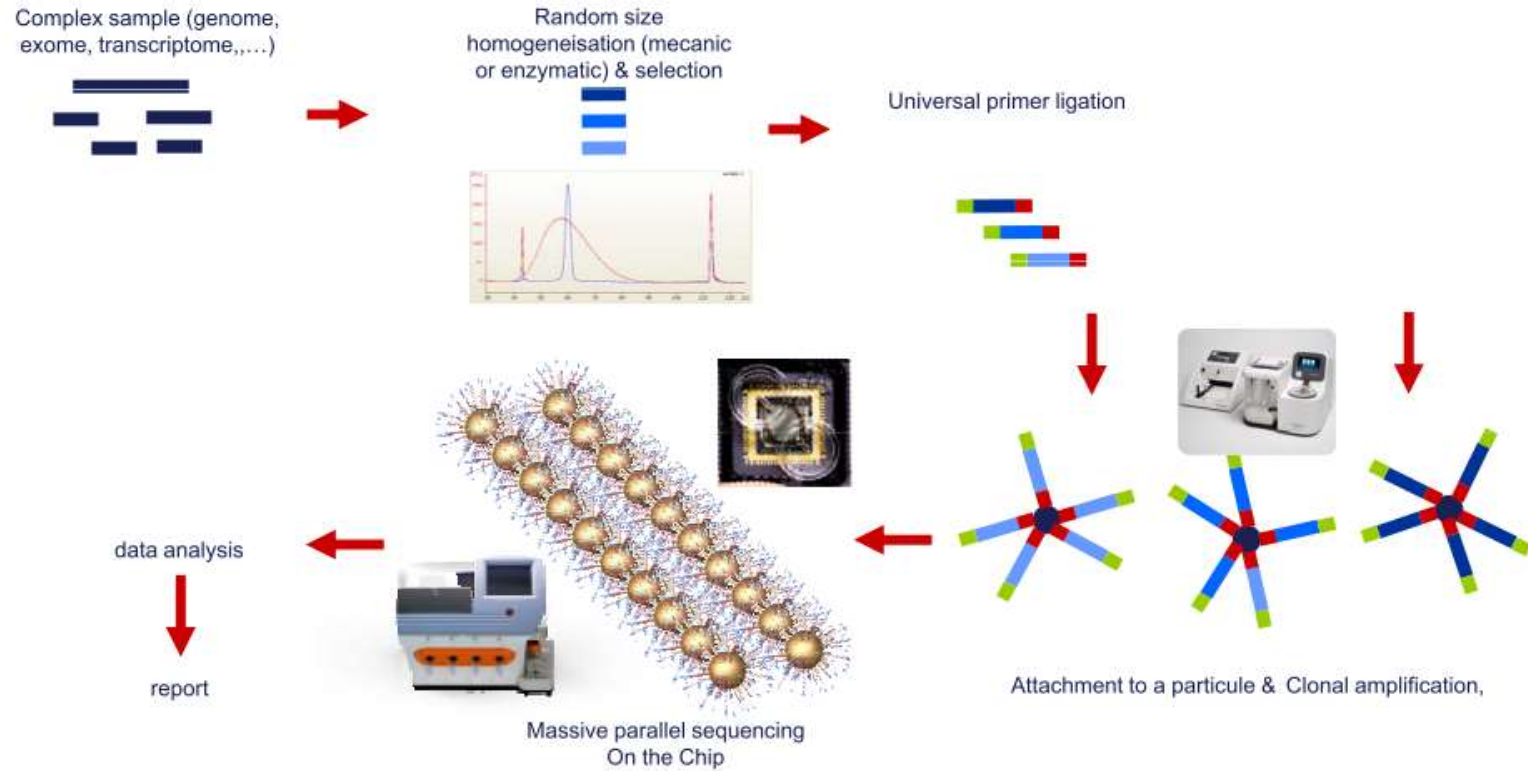
- Χρησιμοποιεί τη χημεία της πυροαλληλούχισης
- DNA πολυμεράση → καταλύει αντίδραση ενσωμάτωσης νουκλεοτιδίων → εκλύεται φώς από το ένζυμο λουσιφεράση
- Ποσότητα φωτός/ανάλογη με τον αριθμό των νουκλεοτιδίων που ενσωματώνονται.

Βιβλιοθήκη θραυσμάτων αναμιγνύεται με σφαιρίδια τα οποία έχουν προσδεδμένα ολιγονουκλεοτίδια συμπληρωματικά με τις αλληλουχίες προσαρμοστών που έχουν χρησιμοποιηθεί για τη δημιουργία της βιβλιοθήκης θραυσμάτων.

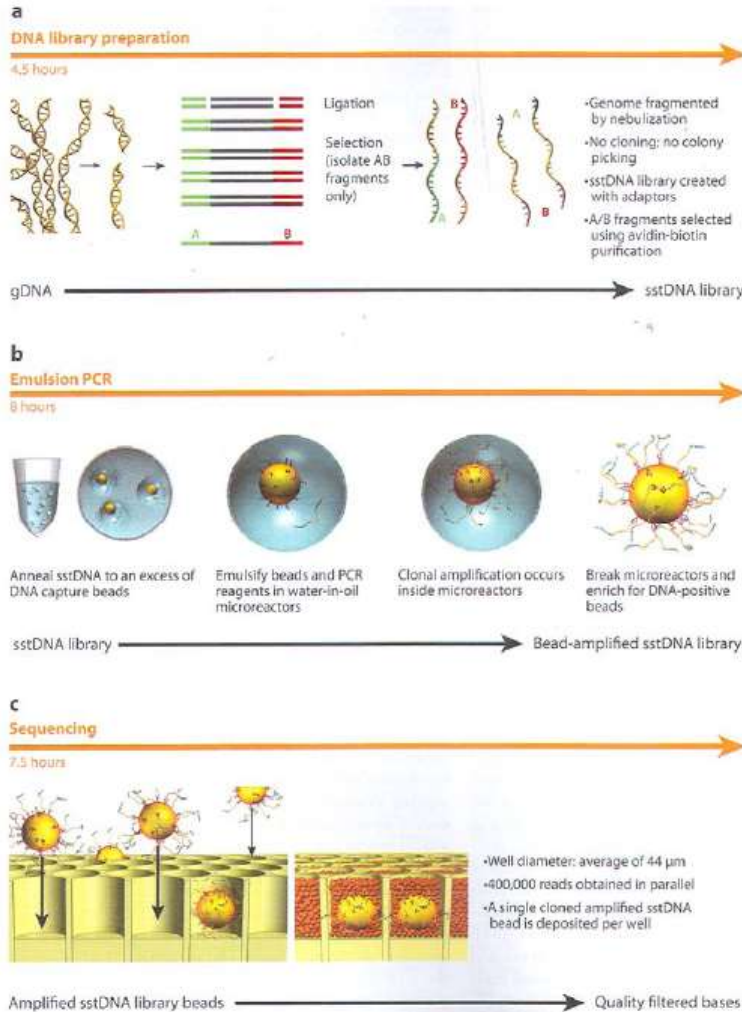
Κάθε σφαιρίδιο προσδένει ένα θραύσμα και απομονώνεται σε ένα υδατικό σταγονίδιο σε λάδι όπου περιέχονται τα αντιδραστήρια της PCR. Η ενίσχυση γίνεται στο μίγμα και παράγονται περίπου 1.000.000 αντίγραφα κάθε θραυσματος σε κάθε σφαιρίδιο.

- Αυτή η διαδικασία ονομάζεται emulsion PCR (PCR σε γαλάκτωμα)
- Τα ενισχυμένα θραύσματα αλληλουχίζονται μαζικά
- Κάθε σφαιρίδιο εισέρχεται σε ένα πηγαδάκι ενός ειδικού πιάτου και προστίθενται σφαιρίδια με τα αντιδραστήρια για την αντίδραση πυροαλληλούχισης.
- Μια CCD camera καταγράφει το φως που εκλύεται από κάθε σφαιρίδιο.
- Αλγόριθμος που επεξεργάζεται τα δεδομένα και παράγει την αλληλουχία
- Τα δεδομένα συλλέγονται σε 'flow space' string of bases + αριθμός βάσεων για κάθε θέση βάσης
- Λάθη σε μεγάλες επαναλαμβανόμενες αλληλουχίες μονονουκλεοτιδίων
- Illumina → sequence space, λάθη στο τέλος του διαβάσματος
- SOLID → 'color space' alignment

Massive Parallel Sequencing workflow



Roche/454 FLX Pyrosequencer



Library fragments are mixed with agarose beads with oligos complementary to adapter sequences on the library.

Each bead is associated with a single fragment.

Each fragment-bead complex is isolated into individual oil:water micelles with PCR mixture.

Thermal cycling of this emulsion PCR of the micelles produces amplified unique sequences on the bead surface.

“En mass” sequencing of PCR products on picotiter plates (PTP) with single beads in each picowell.

Enzyme/substrate containing beads for the pyrosequencing reaction are added to wells that act as flow cells for addition of individual pure nucleotide solutions. The CCD camera records the light emitted at each bead.

Next Gen SEQUENCING



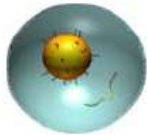
Key Concept

454 Sequencing from individual DNA

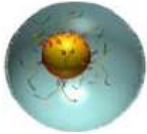
molecules



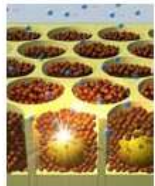
Library of adapted DNA molecules



One DNA molecule per bead



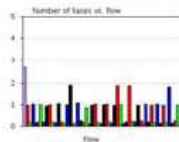
Clonal amplification to ~10 million copies



Independent sequencing of each bead



One Bead = One Read = One DNA molecule



RNA

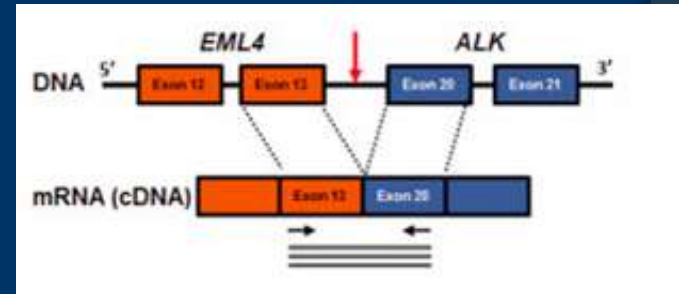


Αντίστροφη μεταγραφή-cDNA

Προετοιμασία Βιβλιοθηκών
PCR με τους εκκινητές του
panel

Ενίσχυση Βιβλιοθηκών με
emulsion PCR

Αλληλούχιση με NGS



DNA vs RNA

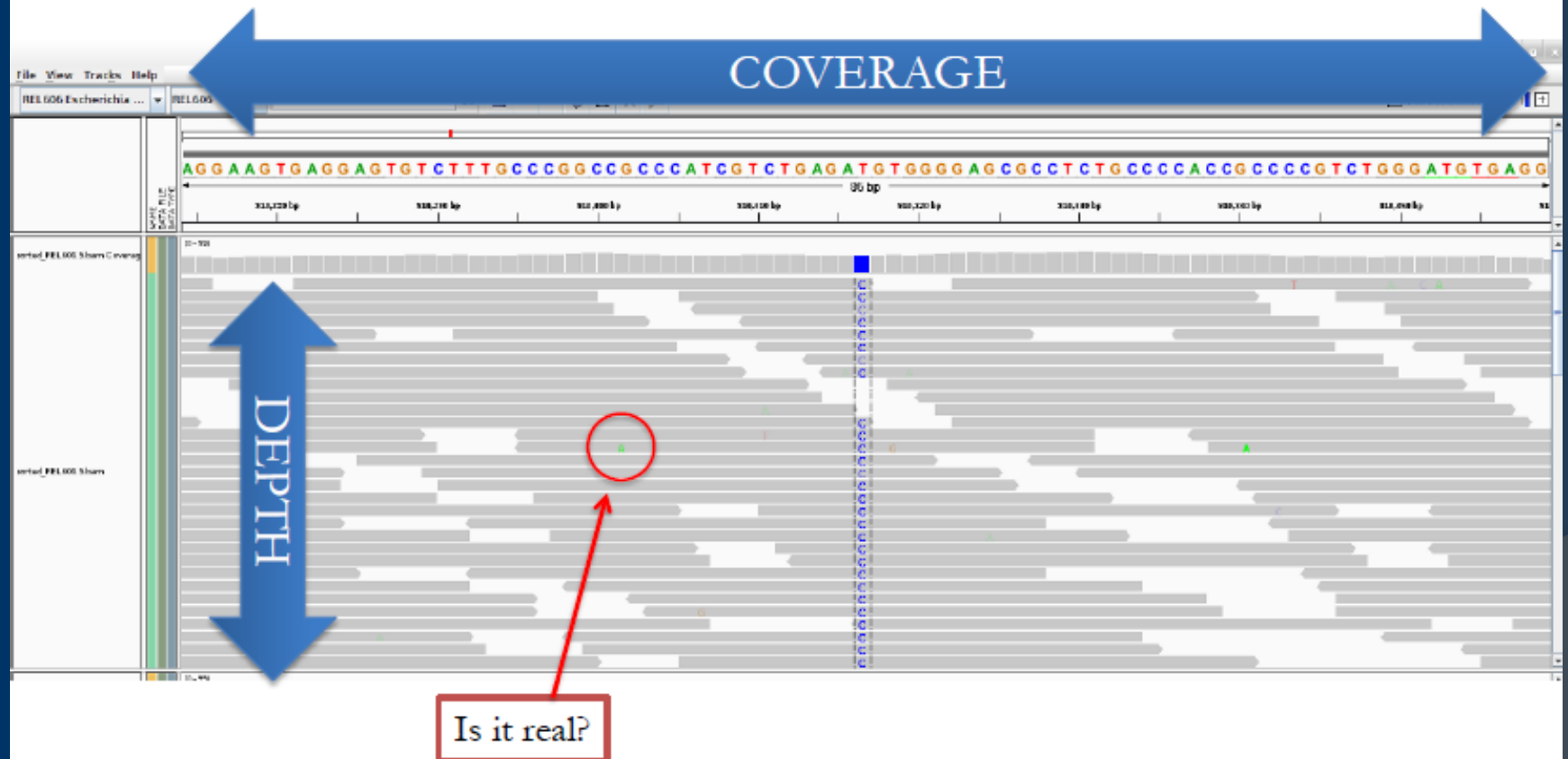
• DNA

- Μεταλλαγές (ελλείψεις, προσθήκες, σημειακές κτλ)
- Αριθμός γονιδιακών αντιγράφων (amplification)

RNA

- Διαμεταθέσεις/Αναδιάταξεις
- Γονίδια σύντηξης (FUSIONS)

Variant Detection



NGS

Coverage recommendations

- De-novo sequencing – 20-30x coverage

- Re-sequencing – 15-25x coverage

- Amplicon sequencing:

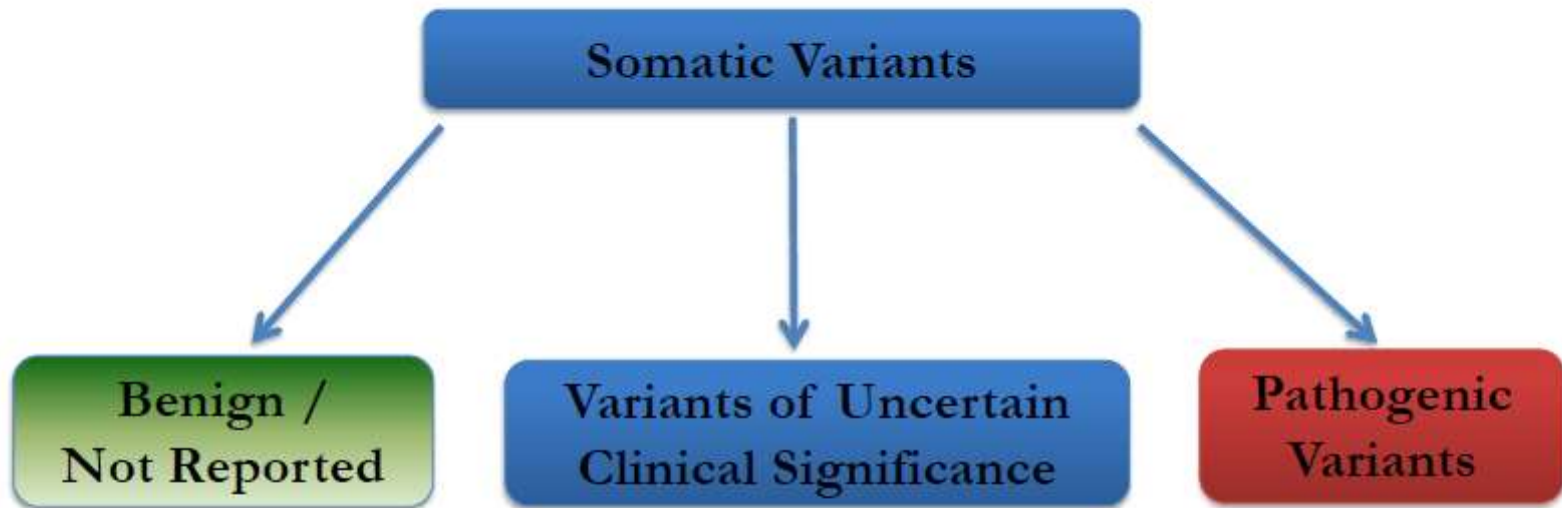
- Heterozygote detection
 - 40x coverage will result in greater than 99.9% chance of detecting a heterozygote
- 5% variation of single base changes or multibase deletions
 - 1000x coverage will average 50 reads for a typical variant
- 1% variation of single base changes or multibase deletions
 - 5000x coverage will average 50 reads for a typical variant
- Single-base indels may require additional depth.

GS Junior System
Guidelines for Amplicon Experimental Design

August 2010

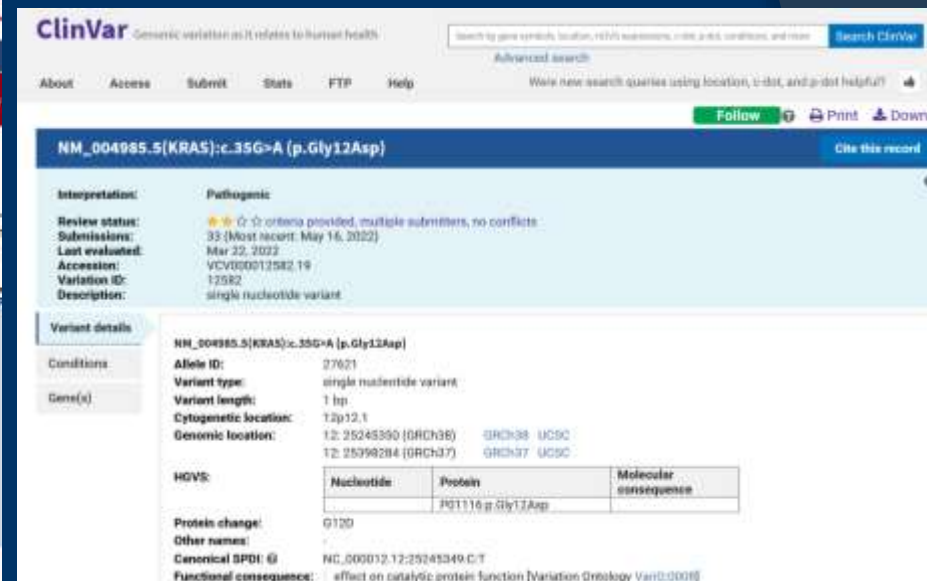
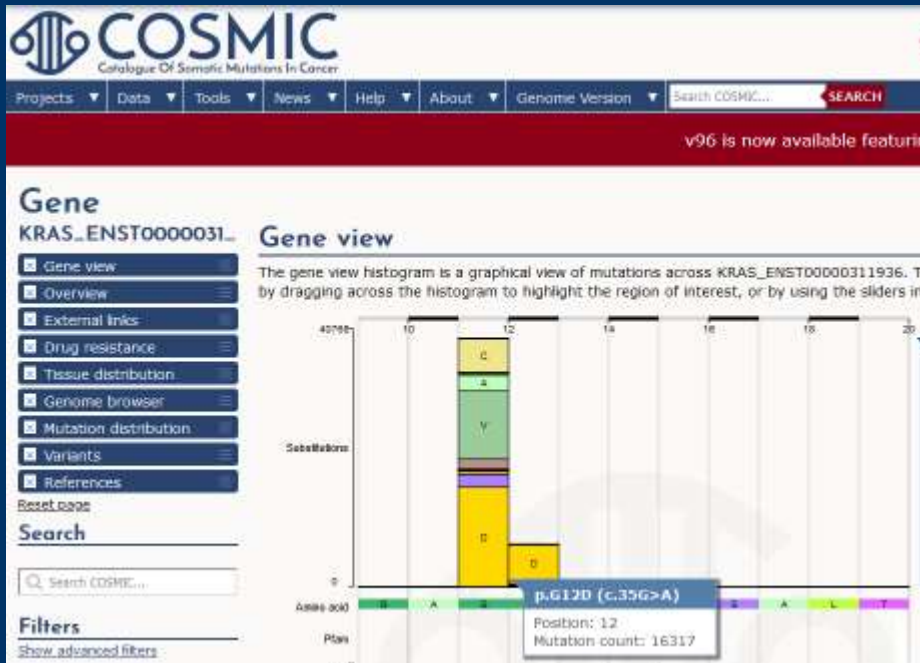
Clinical significance of rare variants?

Variant Classification



ARUP Tiers
Tier 1 – Actionable (FDA Approved Therapies in Patient Tumor Type, Established Diagnostic or Prognostic Significance)
Tier 2 – Potentially Actionable (FDA Approved Therapies in another Tumor Type, Potential Diagnostic or Prognostic Significance)
Tier 3 – Variants of Unknown Significance (VUS)

Ταυτοποίηση της Μεταλλαγής- ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ



Varsome, TP53 database,
GnomAD etc

Mutation

COSV51765161

- Overview ☰
- Tissue distribution ☰
- Samples ☰
- Pathways affected ☰
- References ☰

[Reset page](#)

Overview

This section shows a general overview of the selected mutation. It describes the source of the mutation i.e gene name/sample name/tissue name with unique ID, and also shows sequence level. You can see more information on our [help pages](#).

Genomic Mutation ID ⓘ	COSV51765161
Legacy Identifier ⓘ	COSM6224
Gene name	EGFR
AA mutation	p.L858R (Substitution - Missense, position 858, L→R)
CDS mutation	c.2573T>G (Substitution, position 2573, T→G)
SNP	No
Nucleotides inserted	n/a
Genomic coordinates	GRCh38, 7:55191822..55191822 , view Ensembl contig ↗
CDD	n/a
HomoloGene	n/a
Ever confirmed somatic?	Yes
FATHMM prediction	Pathogenic (score 0.98)
Remark	n/a
Recurrent	n/a
Drug resistance	n/a
Alternative Ids ⓘ	130969846{EGFR_ENST00000454757} , 127653783{EGFR_ENST00000455089} , 168917776{EGFR_ENST00000638463}


 Search ClinVar

Advanced search

[Follow](#) Download

NM_005228.5(EGFR):c.2573T>G (p.Leu858Arg)

[Cite this record](#)

FEEDBACK

Interpretation: drug response

Review status: ★★☆☆ reviewed by expert panel

Submissions: 11 (Most recent: Oct 22, 2019)

Last evaluated: Mar 21, 2018

Accession: VCV000016609.4

Variation ID: 16609

Description: single nucleotide variant

Variant details

Conditions

Gene(s)

NM_005228.5(EGFR):c.2573T>G (p.Leu858Arg)

Allele ID: 31648

Variant type: single nucleotide variant

Variant length: 1 bp

Cytogenetic location: 7p11.2

Genomic location: 7: 55191822 (GRCh38) [GRCh38](#) [UCSC](#)
7: 55259515 (GRCh37) [GRCh37](#) [UCSC](#)

HGVS:

Nucleotide	Protein	Molecular consequence
NC_000007.13:g.55259515T>G		
NC_000007.14:g.55191822T>G		
NM_005228.5:c.2573T>G MANE SELECT	NP_005219.2:p.Leu858Arg	missense

... more HGVS

Protein change: L858R, L591R, L805R, L813R

Other names: -

Canonical SPDI: [NC_000007.14:55191821:T:G](#)

Functional consequence: -

Global minor allele frequency (GMAF): -

Allele frequency: -

Links: UniProtKB: P00533#VAR_019298
OMIM: 131550.0002
dbSNP: rs121434568
PharmGKB Clinical Annotation: 981420042
PharmGKB Clinical Annotation: 981475838
PharmGKB Clinical Annotation: 981475880
[ClinGen: CG126712](#)

NGS tumor panels

Using a standard solid tumor NGS panel we are able to simultaneously analyze mutations in more than 40 genes, including many of the common aberrations seen in lung cancer.

For example, it could provide information on KRAS, BRAF, NRAS, MET, PTEN, RET status, which have emerged as a potential biomarkers for clinical trials

Using NGS facilitates the accumulation of large databases of potential driver mutations that could be useful for future research.

ABL1	EGFR	GNAS	MLH1	RET
AKT1	ERBB2	HNF1A	MPL	SMAD4
ALK	ERBB4	HRAS	NOTCH1	SMARCB1
APC	FBXW7	IDH1	NPM1	SMO
ATM	FGFR1	JAK2	NRAS	SRC
BRAF	FGFR2	JAK3	PDGFRA	STK11
CDH1	FGFR3	KDR	PIK3CA	TP53
CDKN2A	FLT3	KIT	PTEN	VHL
CSF1R	GNA11	KRAS	PTPN11	
CTNNB1	GNAQ	MET	RB1	

CATEGORIZED BY SOMATIC
ALTERATION TYPE

52
GENES

CATEGORIZED BY
PUBLISHED RELEVANCE

Hotspot mutations

AKT1, ALK, AR, BRAF, CDK4, CTNNB1, DDR2, EGFR, ERBB2, ERBB3, ERBB4, ESR1, FGFR2, FGFR3, GNA11, GNAQ, HRAS, IDH1, IDH2, JAK1, JAK2, JAK3, KIT, KRAS, MAP2K1, MAP2K2, MET, MTOR, NRAS, PDGFRA, PIK3CA, RAF1, RET, ROS1, SMO

Focal CNV gains

ALK, AR, BRAF, CCND1, CDK4, CDK6, EGFR, ERBB2, FGFR1, FGFR2, FGFR3, FGFR4, KIT, KRAS, MET, MYC, MYCN, PDGFRA, PIK3CA

Fusion drivers

ABL1, AKT3, ALK, AXL, BRAF, EGFR, ERBB2, ERG, ETV1, ETV4, ETV5, FGFR1, FGFR2, FGFR3, MET, NTRK1, NTRK2, NTRK3, PDGFRA, PPARG, RAF1, RET, ROS1

Labels

ALK, BRAF, EGFR, ERBB2, KRAS, NRAS

Guidelines

ALK, BRAF, EGFR, ERBB2, KIT, KRAS, MET, NRAS, PDGFRA, RET, ROS1

Drug targets

ABL1, AKT1, AKT3, ALK, AR, AXL, BRAF, CCND1, CDK4, CDK6, CTNNB1, DDR2, EGFR, ERBB2, ERBB3, ERBB4, ESR1, FGFR1, FGFR2, FGFR3, FGFR4, IDH1, IDH2, JAK1, JAK2, JAK3, KIT, KRAS, MAP2K1, MAP2K2, MET, MTOR, MYC, NTRK1, NTRK2, NTRK3, PDGFRA, PIK3CA, PPARG, RAF1, RET, ROS1, SMO

NGS

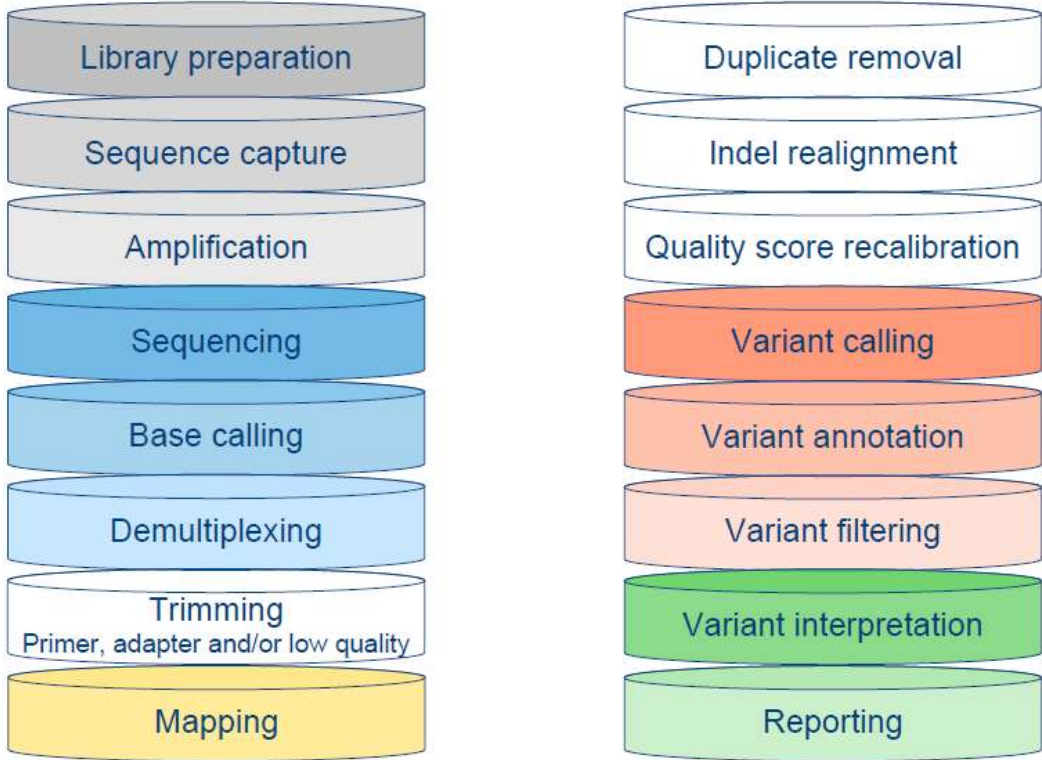
⦿ Advantages

- Very powerful tool/ enormous potential
- Biomarker discovery
- Information on clonality, heterogeneity

⦿ Current Difficulties

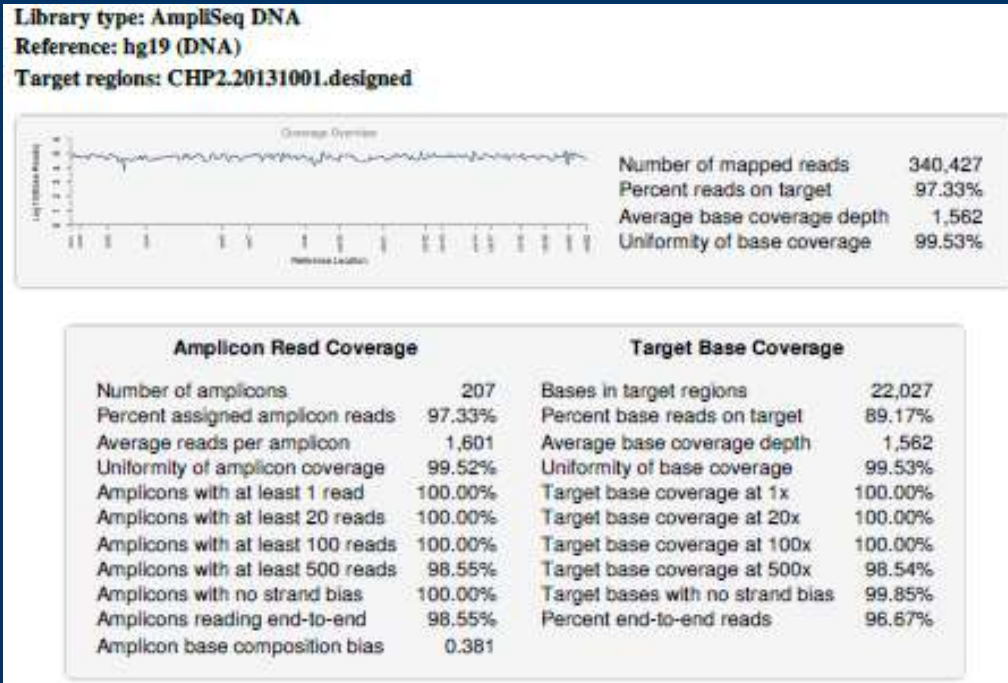
- Enormous data -how to interpret
- Whole framework has to emerge
- Validation, verification
- Low level mutations—importance?
- Clinical value rather limited to druggable driver point mutations
- Cost consideration

NGS workflow

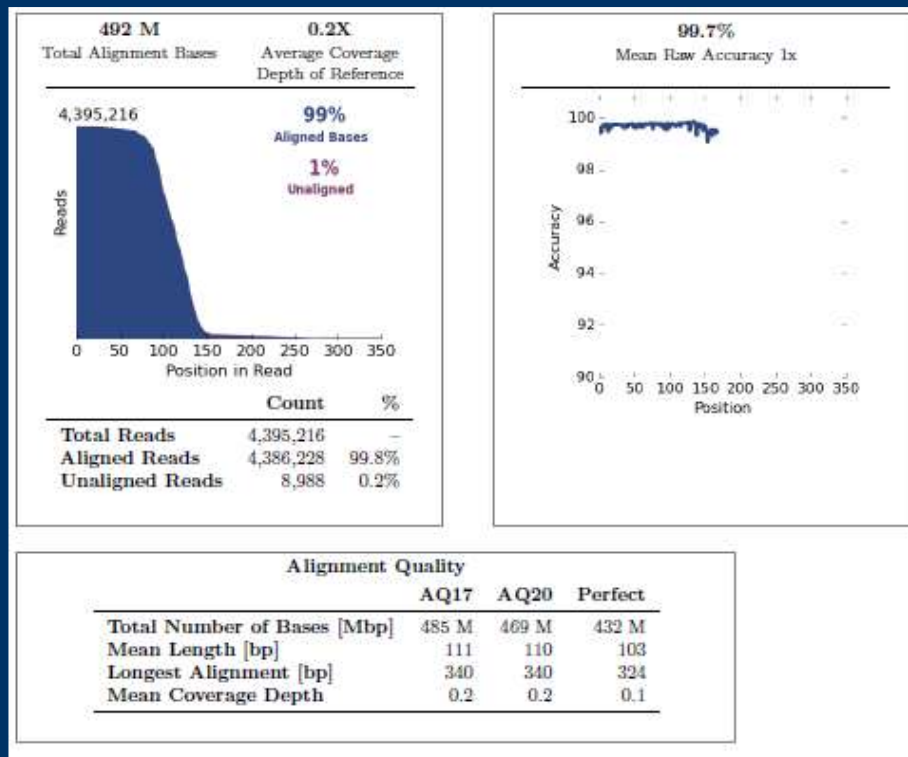


Extensive quality assessment required in all steps/ validation
Diagnostic yield/clinical utility

Παράδειγμα Coverage Analysis



Παράδειγμα Alignment Summary



Αρχικά Κεντρική Εισαγωγή Διάταξη σελίδας Τύπος Διάγραμμα Αναβάθμιση Προβολή

Ανοικτή Ανταλλαγή Πεδίο μορφοποίησης

Μορφοποίηση κελιού Αναδίπλωση κελιού Ζωγραφιστική και εστίαση στα κέντρα

Γραμμοστάσιο Στοιχισμός

Εξασθενήστε Μορφοποίηση από όρους Μορφοποίηση με πλαίσιο

Κανονικό Κακό Καλό Ουδέτερο

Εισαγωγή Διαγραφή Μορφοποίηση

Αυτόματη Αθροιστική Συμπλήρωση Αποκλιση

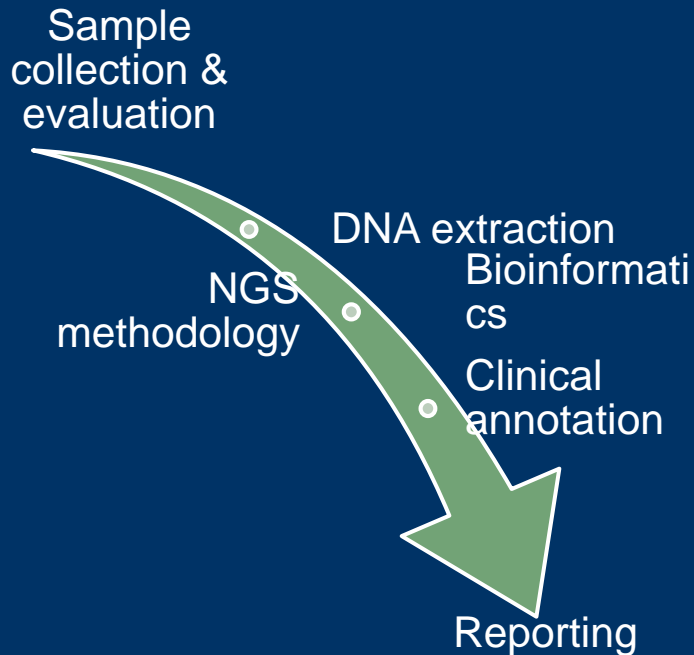
Ταξινόμηση φθίνουσα Ταξινόμηση φθίνουσα

#	locus	type	ref	length	genotype	filter	pvalue	cnv_pve	coverage	allele_c	allele_r	%_freq	maf	hrun	gene	transcrip	location	function	codon	exon	protein
1	##reference=hg19																				
2	##referenceURI=hg19																				
4	chr1:4381	REF	G		G/G	PASS	0.002036807568069		856	G=855,A=-	G=0.9988,0.12			1	MPL	NM_0053	MPL:exonic:NM_005373.2			10	
5	chr1:4381	REF	G		G/G	PASS	1,83E+12		856	G=856,A=C	G=1.0,A=C 0.00			1	MPL	NM_0053	MPL:exonic:NM_005373.2			10	
6	chr1:4381	REF	TGG		TGG/TGG	PASS	1,87E+12		854	TGG=854,	TGG=1.0,AAA=0.00,AAG=0.00,1,1,1,1,1			1	MPL	NM_0053	MPL:exonic:NM_005373.2			10	
7	chr1:4381	REF	G		G/G	PASS	1,81E+12		857	G=857,C=C	G=1.0,C=C 0.00,T=0.00			2,2	MPL	NM_0053	MPL:exonic:NM_005373.2			10	
8	chr1:4381	REF	G		G/G	PASS	0.002112613253436		852	G=851,A=-	G=0.9988,0.12			1	MPL	NM_0053	MPL:exonic:NM_005373.2			10	
9	chr1:1152	REF	C		C/C	PASS	1,85E+07		2000	C=2000,T=C	C=1.0,T=0 0.00			1	NRAS	NM_0025	NRAS:exonic:NM_002524.4			4	
10	chr1:1152	REF	T		T/T	PASS	1,89E+06		1998	T=1998,A=T	A=1.0,A=0 0.00			1	NRAS	NM_0025	NRAS:exonic:NM_002524.4			3	
11	chr1:1152	REF	TTG		TTG/TTG	PASS	4,67E+08		1995	TTG=1993	TTG=0.99,ATG=0.00,CAG=0.00,2,2,2,2,2			2,2,2,2,2	NRAS	NM_0025	NRAS:exonic:NM_002524.4			3	
12	chr1:1152	REF	TG		TG/TG	PASS	7,09E+10		1999	TG=1994,	TG=0.997,AG=0.00,CG=0.25,GC 2,2,2,2,2			2,2,2,2,2	NRAS	NM_0025	NRAS:exonic:NM_002524.4			3	
13	chr1:1152	REF	GT		GT/GT	PASS	1,94E+06		1995	GT=1995,	GT=1.0,CT CT=0.00,TT=0.00,TA= 1,1,1			1,1,1	NRAS	NM_0025	NRAS:exonic:NM_002524.4			3	
14	chr1:1152	REF	C		C/C	PASS	4,56E+07		2000	C=1999,T=C	C=0.9995,0.05			2	NRAS	NM_0025	NRAS:exonic:NM_002524.4			3	
15	chr1:1152	REF	C		C/C	PASS	1,85E+07		2000	C=2000,T=C	C=1.0,T=0 0.00	0.0		1	NRAS CSE	NM_0025	NRAS:exonic:NM_0025			2	
16	chr1:1152	REF	AC		AC/AC	PASS	4,53E+08		1998	AC=1996,	AC=0.999,GA=0.00,GC=0.10			1,1	NRAS CSE	NM_0025	NRAS:exonic:NM_0025			2	
17	chr1:1152	REF	C		C/C	PASS	4,45E+09		2000	C=1998,A=C	C=0.999,A A=0.00,G=0.00,T=0.12,2,2			2,2,2	NRAS CSE	NM_0025	NRAS:exonic:NM_0025			2	
18	chr1:1152	REF	C		C/C	PASS	1,65E+10		2000	C=1996,A=C	C=0.998,A A=0.00,G=0.00,T=0.22,2,2			2,2,2	NRAS CSE	NM_0025	NRAS:exonic:NM_0025			2	
19	chr1:1152	REF	A		A/A	PASS	6,86E+10		2000	A=1997,G=A	A=0.9985,0.15			1	NRAS CSE	NM_0025	NRAS:exonic:NM_0025			2	
20	chr1:1152	REF	CC		CC/CC	PASS	2,58E+12		1996	CC=1990,	CC=0.997,AC=0.00,GC=0.00,TC 2,2,2,2			2,2,2,2	NRAS CSE	NM_0025	NRAS:exonic:NM_0025			2	
21	chr1:1152	REF	C		C/C	PASS	1,65E+10		2000	C=1996,A=C	C=0.998,A A=0.00,G=0.00,T=0.22,2,2			2,2,2	NRAS CSE	NM_0025	NRAS:exonic:NM_0025			2	
22	chr1:1152	REF	C		C/C	PASS	4,56E+07		2000	C=1999,T=C	C=0.9995,0.05			1	NRAS CSE	NM_0025	NRAS:exonic:NM_0025			2	
23	chr2:2943	REF	C		C/C	PASS	1,57E+11		1100	C=1100,T=C	C=1.0,T=0 0.00			1	ALK	NM_0043	ALK:exonic:NM_004304.4				25
24	chr2:2944	REF	G		G/G	PASS	1,09E+11		1137	G=1137,T=G	G=1.0,T=0 0.00			1	ALK	NM_0043	ALK:exonic:NM_004304.4				23
25	chr2:2944	REF	G		G/G	PASS	9,65E+09		1149	G=1149,C=G	G=1.0,C=C 0.00,T=0.00			1,1	ALK	NM_0043	ALK:exonic:NM_004304.4				23
26	chr2:2944	REF	A		A/A	PASS	9,75E+09		1148	A=1148,C=A	A=1.0,C=0 0.00			2	ALK	NM_0043	ALK:exonic:NM_004304.4				23
27	chr2:2944	REF	A		A/A	PASS	9,75E+09		1148	A=1148,C=A	A=1.0,C=0 0.00,G=0.00,T=0.02,2,2			2,2,2	ALK	NM_0043	ALK:exonic:NM_004304.4				23
28	chr2:2091	REF	G		G/G	PASS	1,87E+07		1999	G=1999,T=G	G=1.0,T=0 0.00			1	IDH1	NM_0058	IDH1:exonic:NM_005896.3				4
29	chr2:2091	REF	A		A/A	PASS	1,90E+06		1997	A=1997,T=A	A=1.0,T=0 0.00			1	IDH1	NM_0058	IDH1:exonic:NM_005896.3				4
30	chr2:2091	REF	AC		AC/AC	PASS	2,06E+06		1989	AC=1989,	AC=1.0,G 0.00			1	IDH1	NM_0058	IDH1:exonic:NM_005896.3				4
31	chr2:2091	REF	CG		CG/CG	PASS	7,82E+11		1997	CG=1990,	CG=0.996,AG=0.00,TG=0.35,AC 1,1,1			1,1,1	IDH1	NM_0058	IDH1:exonic:NM_005896.3				4
32	chr2:2091	REF	G		G/G	PASS	3,12E+08		2000	G=1997,A=G	G=0.9985,A=0.10,C=0.00,T=0.01,1,1,1			1,1,1	IDH1	NM_0058	IDH1:exonic:NM_005896.3				4
33	chr2:2091	REF	T		T/T	PASS	3,13E+08		2000	T=1997,C=T	C=0.9985,0.15			1	IDH1	NM_0058	IDH1:exonic:NM_005896.3				4
34	chr2:2091	REF	C		C/C	PASS	4,45E+08		2000	C=1998,T=C	C=0.999,T 0.10			2	IDH1	NM_0058	IDH1:exonic:NM_005896.3				4

Collection of high redundancy data

- ◎ Next generation sequencing παράγει εκατομμύρια αλληλουχίες και μπορεί να ανιχνεύσει σπάνιες αλληλουχίες που δεν μπορούν να αποκαλυφθούν με άλλες διαδικασίες
- ◎ Στην έρευνα για τον καρκίνο μπορούμε με αυτόν τον τρόπο να ανιχνεύσουμε σπάνια αλληλόμορφα (sequence variants) σε πληθυσμούς καρκινικών κυττάρων
- ◎ Το next generation sequencing είναι ιδιαίτερα ελκυστικό διότι λόγω του μεγάλου αριθμού αλληλουχιών που παράγονται τα αποτελέσματα έχουν μεγαλύτερη στατιστική σημασία.

Procedure



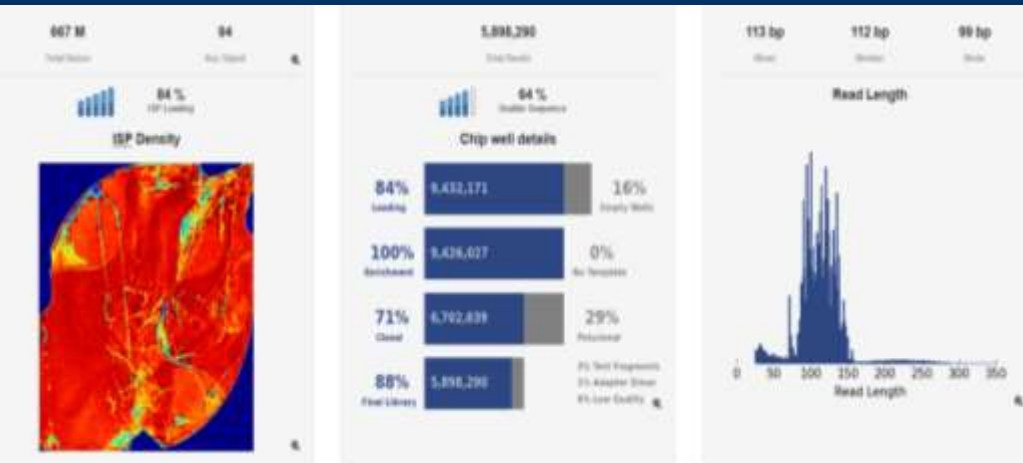
Methodology: NGS/ Ion Ampiseq custom panel

Platform: Ion PGM System

Bioinformatics pipeline: Torrent Suite (Variant Caller/Ion Reporter)



NGS results



Bases	#>200 Bases	Reads	Mean Read Length	Read Length Histogram
7,574,507	7,701,276	63,381	119 bp	
48,730,504	49,897,502	424,024	107 bp	
62,737,529	66,169,600	529,217	120 bp	
35,591,853	34,069,278	303,439	117 bp	
286,304	249,011	2,210	120 bp	
48,495,380	44,878,300	412,881	112 bp	
78,951,027	74,848,981	675,793	115 bp	
66,443,196	67,890,632	525,412	115 bp	
756,838	723,347	18,624	71 bp	
48,893,740	44,775,660	378,393	123 bp	
11,727,481	15,288,588	192,338	104 bp	
35,293,244	33,095,378	271,040	129 bp	
7,100,425	6,713,811	54,474	130 bp	
39,034,471	28,329,878	268,621	117 bp	
42,634,873	41,032,794	383,094	108 bp	
22,821,888	21,976,124	218,429	104 bp	
3,821,454	3,689,942	38,749	88 bp	

NGS coverage analysis report

Mapped Reads	On Target	Mean Depth	Uniformity
20,316	0.20%	0.342	94.86%
362,581	97.37%	3,703	98.91%
363,057	97.16%	3,763	100.00%
355,869	98.52%	3,741	99.99%
269,992	98.26%	2,734	98.65%
120,058	91.54%	1,110	98.65%
526,320	98.15%	2,425	99.48%
12,748	68.56%	78.55	20.38%
827,311	62.80%	5,154	100.00%
517,524	97.64%	5,230	100.00%
123,979	98.28%	1,288	93.04%
480,101	97.70%	4,689	98.65%
318,041	5.57%	42.44	0.63%
426,892	12.78%	126.4	0.43%
406,592	8.18%	82.99	0.46%
70	0.00%	0	100.00%
252,880	3.91%	26.4	0.73%
49,205	5.81%	6.615	0.98%
289,622	0.00%	0	100.00%

Sanger sequencing vs NGS

- ⦿ **Sanger sequencing**: average sequence of all DNA in tube / cannot resolve low frequency mutations.
- ⦿ **NGS** multiple platforms
- ⦿ Radically different underlying chemistries
- ⦿ one read/one input molecule
- ⦿ Every molecule/ individual sequence
- ⦿ Low abundance signals still legible
- ⦿ More depth-rarer mutation

NGS

⦿ Advantages

- Very powerful tool/ enormous potential
- Biomarker discovery
- Information on clonality, heterogeneity

⦿ Current Difficulties

- Enormous data -how to interpret
- Whole framework has to emerge
- Validation, verification
- Low level mutations—importance?
- Clinical value rather limited to druggable driver point mutations
- Cost consideration

Διαμεταθέσεις γονιδίων στη διάγνωση

Tumor Type	Translocation	Gene(s)
Aneurysmal bone cyst	t(16;17)(q22;p13)	CDH11-USP6
	t(1;17)(p34.3;p13)	THRAS1-USP6
	t(3;17)(q21;p13)	CMBP-USP6
	t(9;17)(q22;p13)	DMD-USP6
Angiofibroma	t(17;17)(q21;p13)	COL1A2-USP6
	t(5;8)(p15;q13)	AHRH-NCOA2
Angiomatoid fibrous histiocytoma	t(12;22)(q13;q12)	EWSR1-ATF1
	t(2;22)(q33;q12)	EWSR1-CREB1
	t(12;16)(q13;p11)	FUS-ATF1
Alveolar rhabdomyosarcoma	t(2;13)(q35;q14)	PAK3-FKH1
	t(2;2)(q35;p23)	PAK3-NCOA1
	t(X;2)(q35;q13)	PAK3-AFX
	t(1;13)(p36;q14)	PAK7-FKH1
Alveolar soft-part sarcoma	der(17)t(X;17)(p11;q25)	ASPSR1-TFE3
Chondroid lipoma	t(11;16)(q13;p13)	C11orf95-ANKK2
Clear-cell sarcoma	t(12;22)(q13;q12)	EWSR1-ATF1
	t(2;22)(q33;q12)	EWSR1-CREB1
Congenital/infantile fibrosarcoma	t(12;15)(p13;q25)	ETV6-NTRK3
Dermatofibrosarcoma protuberans/giant cell fibroblastoma	t(17;22)(q21;q13)	COL1A1-PDGFB
Desmoplastic small round-cell tumor	t(11;22)(p13;q12)	EWSR1-WT1
Endometrial stromal sarcoma	t(7;17)(p15;q11)	JAZF1-JAZF1(SU212)
	t(6;7)(p21;p15)	JAZF1-PHF1
	t(8;10)(p21;p11)	EPC-PHF1
	t(10;17)(q22;p13)	FUSHAF-FAM22A/B
	t(1;6)(p32-34;p31)	MEAF6-PHF1

Tumor Type	Translocation	Gene(s)
Epithelioid hemangioendothelioma	t(1;3)(p36;q25)	WWTR1-CAMTA1
	t(X;11)(p11.2;q22.1)	TFE3-YAPI
Ewing sarcoma; Primitive neuroectodermal tumor (PNET)	t(11;12)(q24;q12)	EWSR1-FLI1
	t(21;22)(q22;q12)	EWSR1-ERG
	t(7;22)(p22;q12)	EWSR1-ETV1
	t(2;22)(q33;q12)	EWSR1-FEV
	t(17;22)(q12;q12)	EWSR1-E1AF
	inv(22)(q21;12)	EWSR1-ZSG
Ewing-like bone sarcoma	t(16;21)(p11;q22)	FUS-ERG
	inv(X)(p11.4p11.22)	BCOR-CCNB3
	t(20;22)(q13;q12)	EWSR1-NFATC2
Extraskeletal myxoid chondrosarcoma	t(9;22)(q22;q12)	EWSR1-NR4A3
	t(9;17)(q22;q11)	TAF15-NR4A3
	t(9;15)(q22;q21)	TCF12-NR4A3
Giant cell fibroblastoma	t(3;9)(q11;q22)	TFG-NR4A3
	t(17;22)(q21;q13)	COL1A1-PDGFB
Inflammatory myofibroblastic tumor	t(1;2)(q22;p23)	TPM3-ALK
	t(2;19)(p23;p13)	TPM4-ALK
	t(2;17)(p23;q23)	CLTC-ALK
	t(2;2)(p23;q13)	RANBP2-ALK
	t(2;11)(p23;p15)	CARS-ALK
	inv(2)(p23;q35)	ATIC-ALK
	t(2;4)(p23;q21)	SEC31A-ALK

Διαμεταθέσεις γονιδίων στη διάγνωση

Liposarcoma	t(7;8)(p22;p11)	COL1A2-PL403
Lipoma	t(3;12)(q17;q14-q15)	HMGA2-LPP LPP-HMGA2
Low-grade fibromyxoid sarcoma	t(7;14)(q31;p11)	FUS-CREB1/2
	t(11;16)(q11;p11)	FUS-CREB1
Meningeal hemangiopericytoma	inv(12)(q13q13)	NAB2-STAT6
Monoclonal chondrosarcoma	t(8;8)(q11;q21)	NET1-NCMA2
Myoepithelial tumor of soft tissue and bone	t(8;22)(p21;q12)	EWAS1-PCU093
	t(18;22)(q11;q12)	EWAS1-ZNF444
	t(1;22)(q13;q12)	EWAS1-PBX1
	t with 16p11	FUS
Myoid/round cell sarcoma	t(12;16)(q13;p13)	FUS-GRB7/2
	t(12;22)(q11;q12)	EWAS1-DDIT3
Nuclear Testis	t(17;22)(q13;q13)	MVW9-SPK6
Pelvic sarcoma	t(7;17)(p22;q13)	ACTB-GLI3
Pulmonary myxoid sarcoma	t(2;22)(q14;q12)	EWAS1-CREB1
Sclerosing epithelial fibrosarcoma	t(7;14)(p22;q24)	FUS-CREB1/2
	t(3;13)(q11;q11.2)	FUS-CREB1/2
Solitary fibrous tumor	inv(12)(q13q13)	NAB2-STAT6
Synovial cell rhabdomyosarcoma	t(8;8)(p21;q13)	SRF-NCMA2
	t(8;11)(q13;q13)	TFAD-NCMA2
Synovial sarcoma	t(X;18)(p11;q11)	SS18-SSX1 SS18-SSX2 SS18-SSX4 BCOR-CCNB3 BCOR-MAN1L3 ZC3H7B-BCOR
BCOR-rearranged sarcoma	inv(X)(p11.4q11.22)	
DC-rearranged sarcoma (undifferentiated small round cell/blue cell tumor)	t(4;19)(q25;q11) or t(20;18)(q25;q11)	CIC-DUX4
	t(8;19)(q13;q13.2)	CIC-FORX4

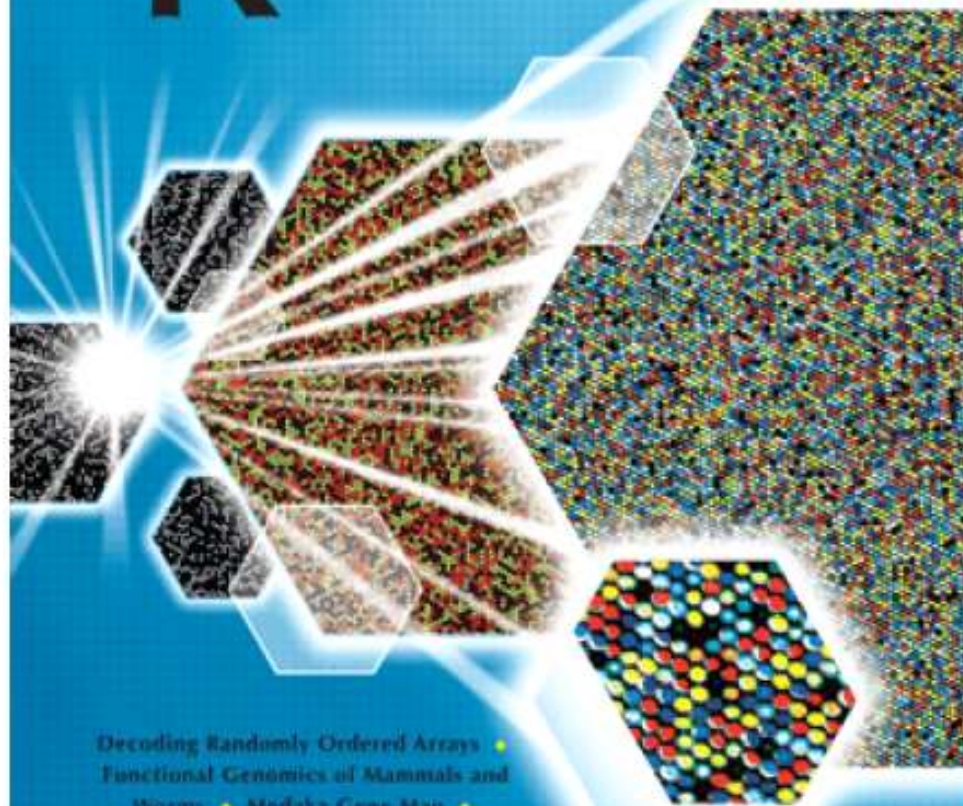
Tumor	Common Fusion Genes
Alveolar Rhabdomyosarcoma	<i>PAX3/7-FOXO1, PAX3-AFX</i>
Angiomatoid Fibrous Histiocytoma	<i>EWAS1-ATF1/CREB1, FUS-ATF1</i>
Alveolar Soft Part Sarcoma	<i>ASPL-TFE3</i>
Clear Cell Sarcoma	<i>EWAS1-ATF1/CREB1</i>
Dermatofibrosarcoma Protuberans	<i>COL1A1-PDGFB</i>
Desmoplastic Small Round Cell Tumor	<i>EWAS1-WT1</i>
Epithelioid Hemangioendothelioma	<i>WWTR1-CAMTA1, YAP1-TFE3</i>
Ewing Sarcoma	<i>EWAS1-FLI1/ERG/FEV/ETV1/E1AF/ZSG, FUS-ERG</i>
Solitary Fibrous Tumor	<i>NAB2-STAT6</i>
Synovial Sarcoma	<i>SS18-SSX1/2/4</i>

ISSN 1088-9051

May 2004

GENOME RESEARCH

Volume 14 Number 5



Decoding Randomly Ordered Arrays •
Functional Genomics of Mammals and
Worms • Medaka Gene Map •
ENSEMBL Special

Cold Spring Harbor
Laboratory Press

