



Summer School 2015

**Alloreattività e trapianti nell'uomo:
le nuove metodiche di studio
e i trapianti alternativi**

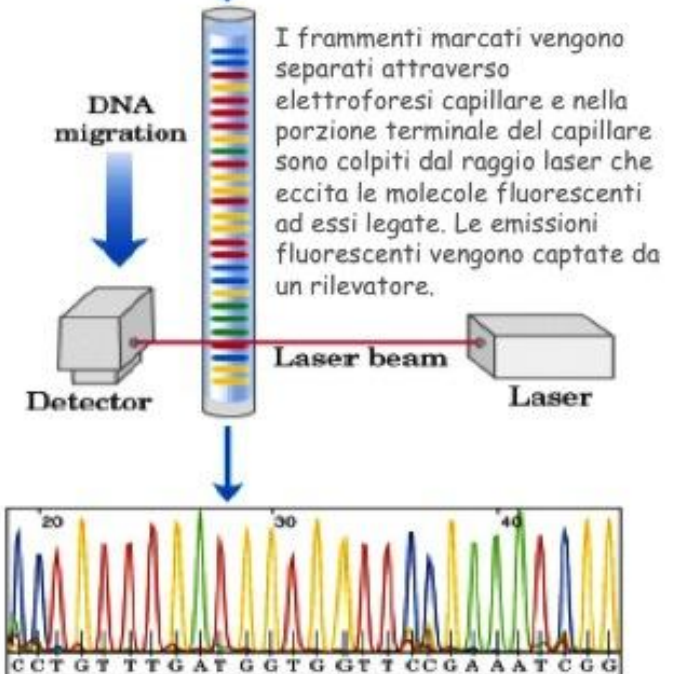
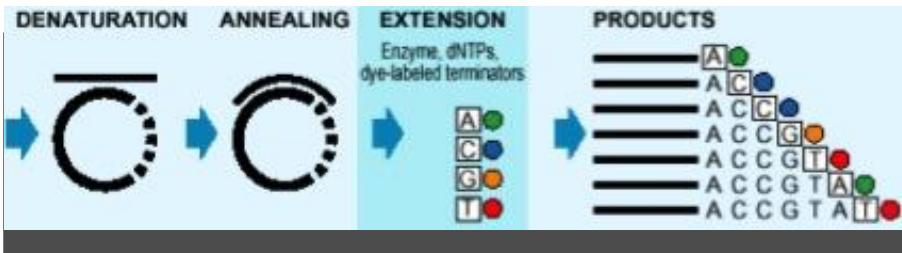
**Next Generation Sequencing :
Principi**

Alessandra Santoro

04 - 06 giugno 2015 Villaggio Cala la Luna Favignana (TP)

Sequenziamento di Sanger

Coniugando a ciascun ddNTP un diverso marcatore fluorescente, è possibile effettuare le quattro reazioni di sequenziamento in un'unica provetta



Le informazioni vengono integrate e trasformate in picchi di colore diverso, con aree proporzionali all'intensità di emissione.



Read di 300-600nt

X

96 read



0.6Mb



Sensibilità 20%

Next generation sequencing

- *Sequenziamento dell'intero genoma di uno o più organismi*
- *Alta velocità di analisi di un'elevata mole di dati*
- *Sensibilità pari all'1%*



30 Mb

Notevoli potenzialità diagnostiche

Diverse technologie

- Ion torrent



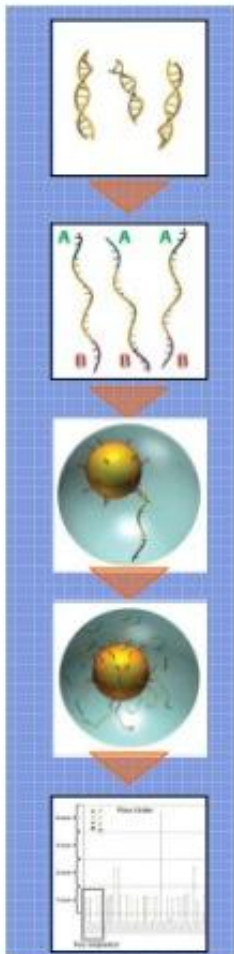
- Genome analyzer Illumina



- 454 Roche technology



Tre fasi...

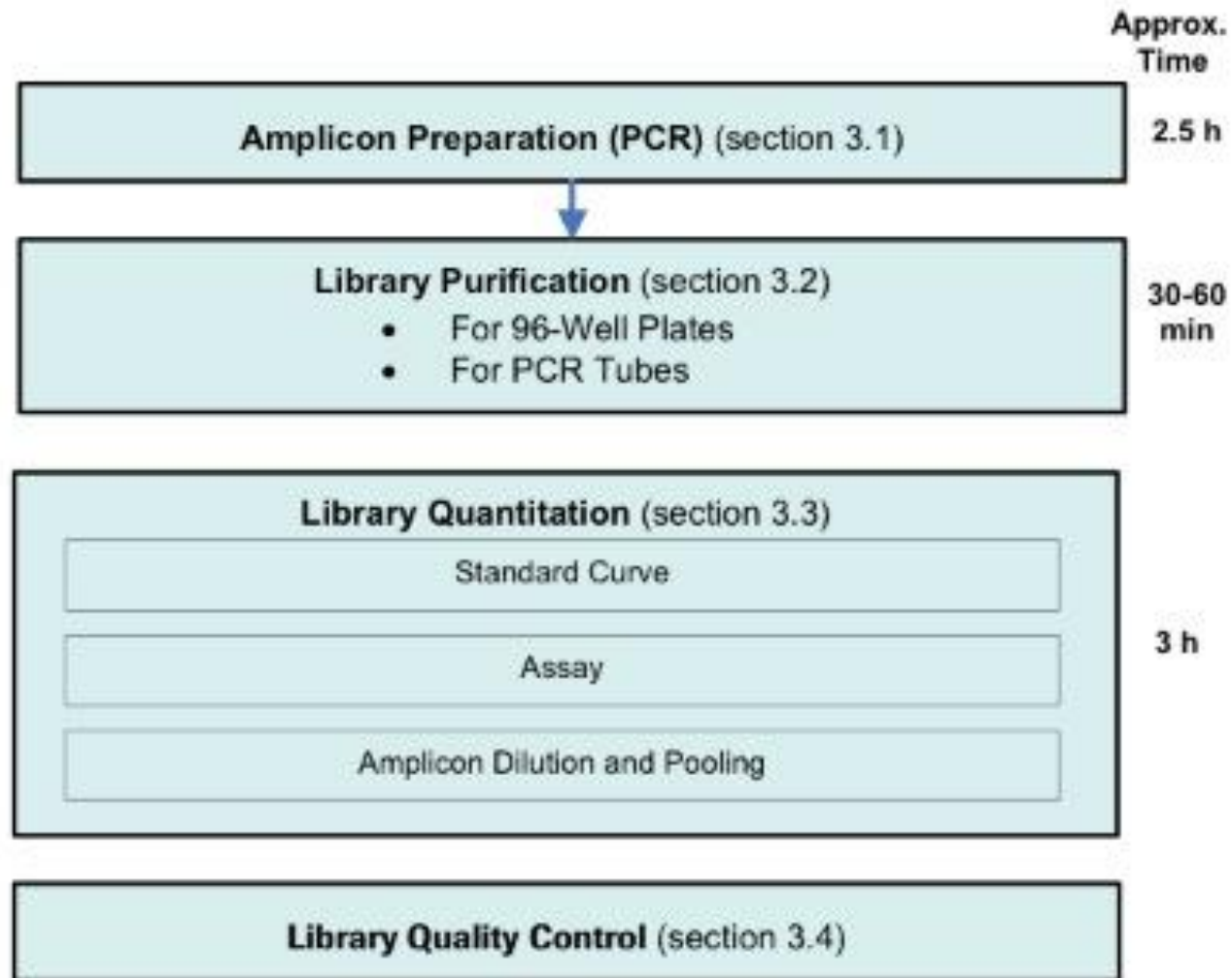


1. Amplicon Library Prep

2. emPCR amplification- Clonal amplification of library fragments

3. Sequencing- Prepare PicoTiterPlate device and start run

1: Ampliconi ➔ Libreria di ampliconi



Molte variabili...

- Numero di ampliconi
- Lunghezza degli ampliconi
- Tipo di sequenziamento (bidirezionale)

Due diversi approcci

GS junior titanium



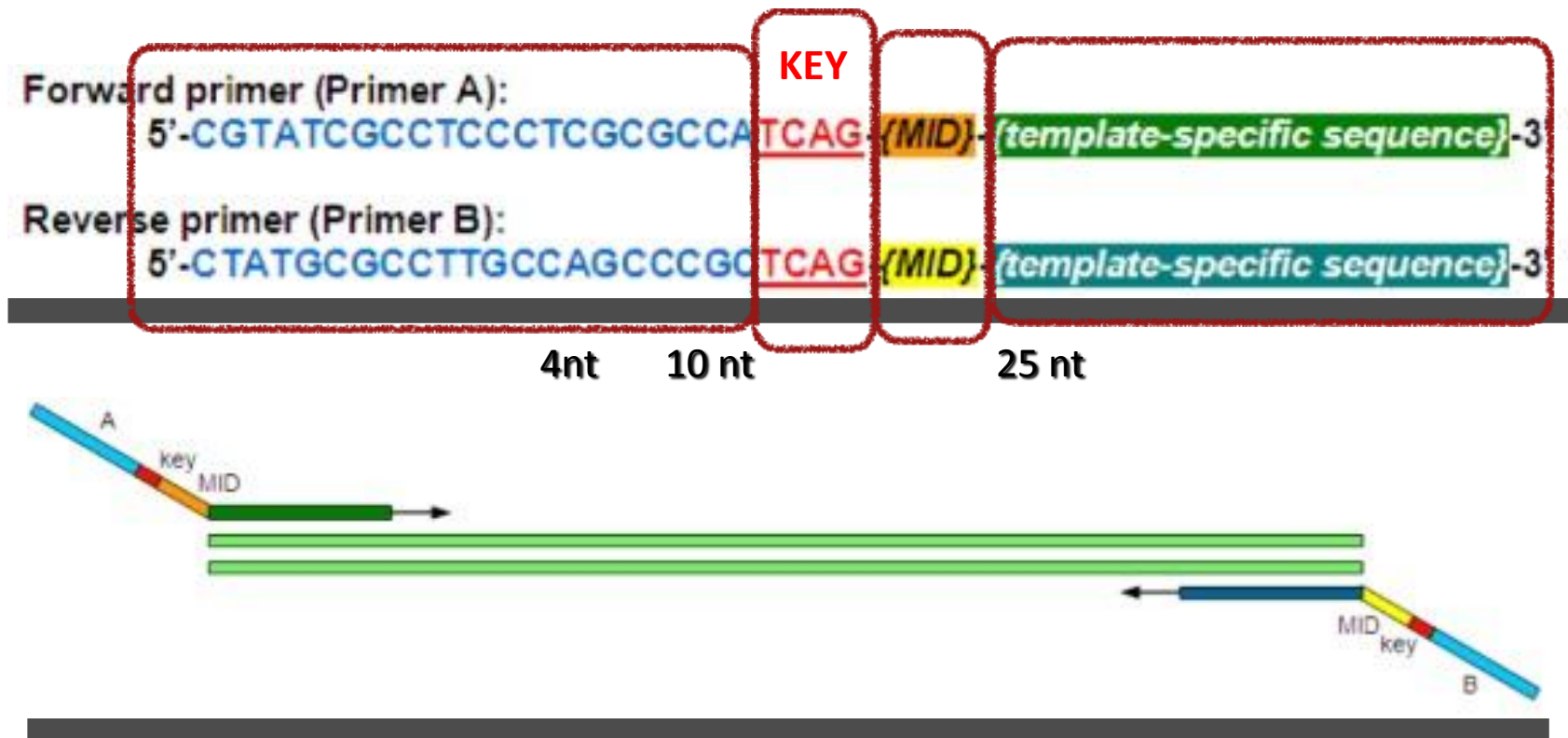
Basic

Universal tailed

BASIC amplicon sequencing



- Semplice
- Dispendioso in termini di tempo e di lavoro



Universal tailed amplicon library preparation



- Tanti campioni o ampliconi
- Riduzione numero di primers

- Parte della sequenza non è informativa

Forward primer (for Primer A side):

5'-[universal tail A (Univ-A)]-[template-specific sequence]-3'

Reverse primer (for Primer B side):

5'-[universal tail B (Univ-B)]-[template-specific sequence]-3'

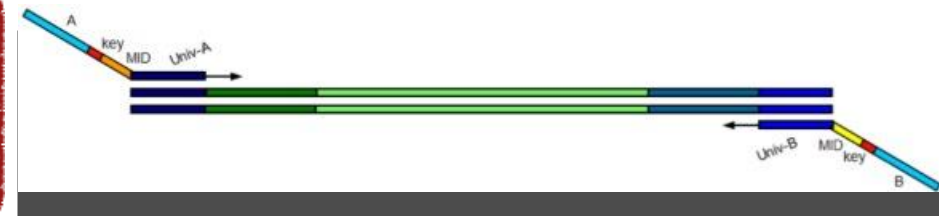


Forward primer (Primer A):

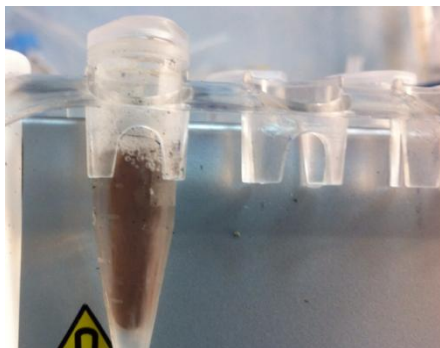
5'-CGTATCGCCTCCCTCGCGCCA TCAG-[MID]-[universal tail A (Univ-A)]-3'

Reverse primer (Primer B):

5'-CTATGCGCCTTGCCAGCCCGC TCAG-[MID]-[universal tail B (Univ-B)]-3'



Purificazione

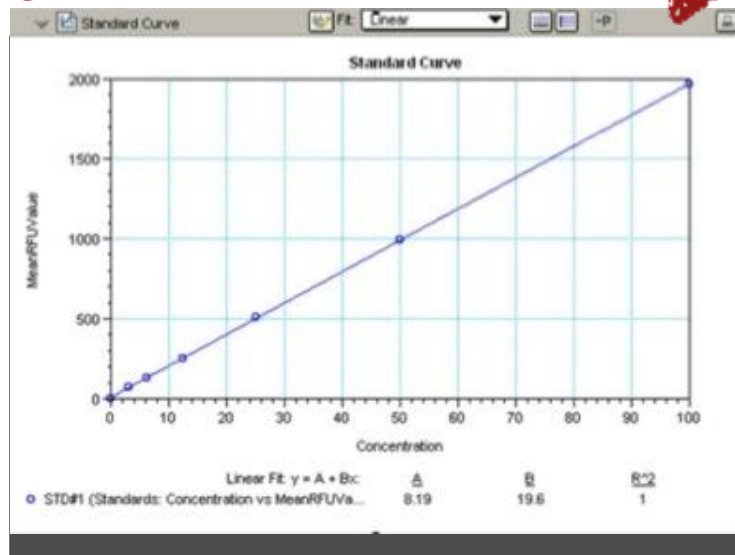


AMPure XP beads magnetiche

Legano in modo aspecifico gli ampliconi

Amplicone : AMPure 1:1 - 1:0.7

Quantizzazione



Quant-it Picogreen 10E9

Tube #	Well	DNA Concentration
Tube 1	A12	100.00ng/well
Tube 2	B12	50.00ng/well
Tube 3	C12	25.00ng/well
Tube 4	D12	12.50ng/well
Tube 5	E12	6.25 ng/well
Tube 6	F12	3.13 ng/well
Tube 7	G12	1.56 ng/well
Tube 8	H12	0.00 ng/well

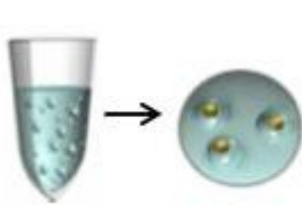
Pooling dei campioni



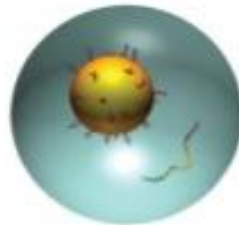
10E6

Molecole DNA / Beads = 1:1,4 - 1:0.8

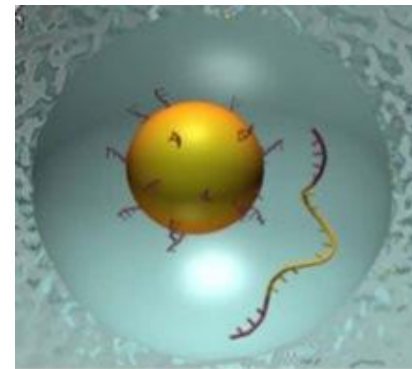
2:Emulsion-PCR amplification



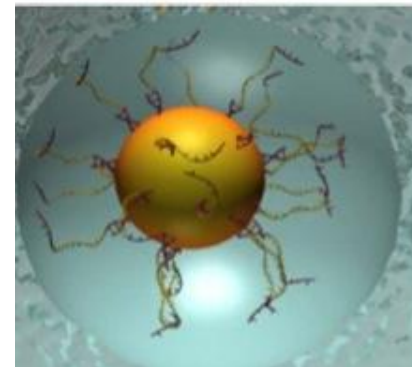
Mix DNA with Capture Beads



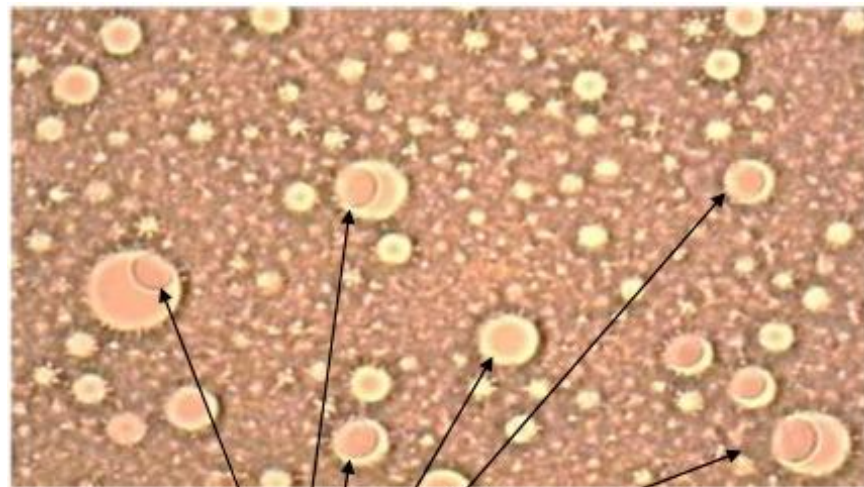
Emulsify DNA Capture Beads and PCR reagents in water-in-oil micoreactors



Before PCR

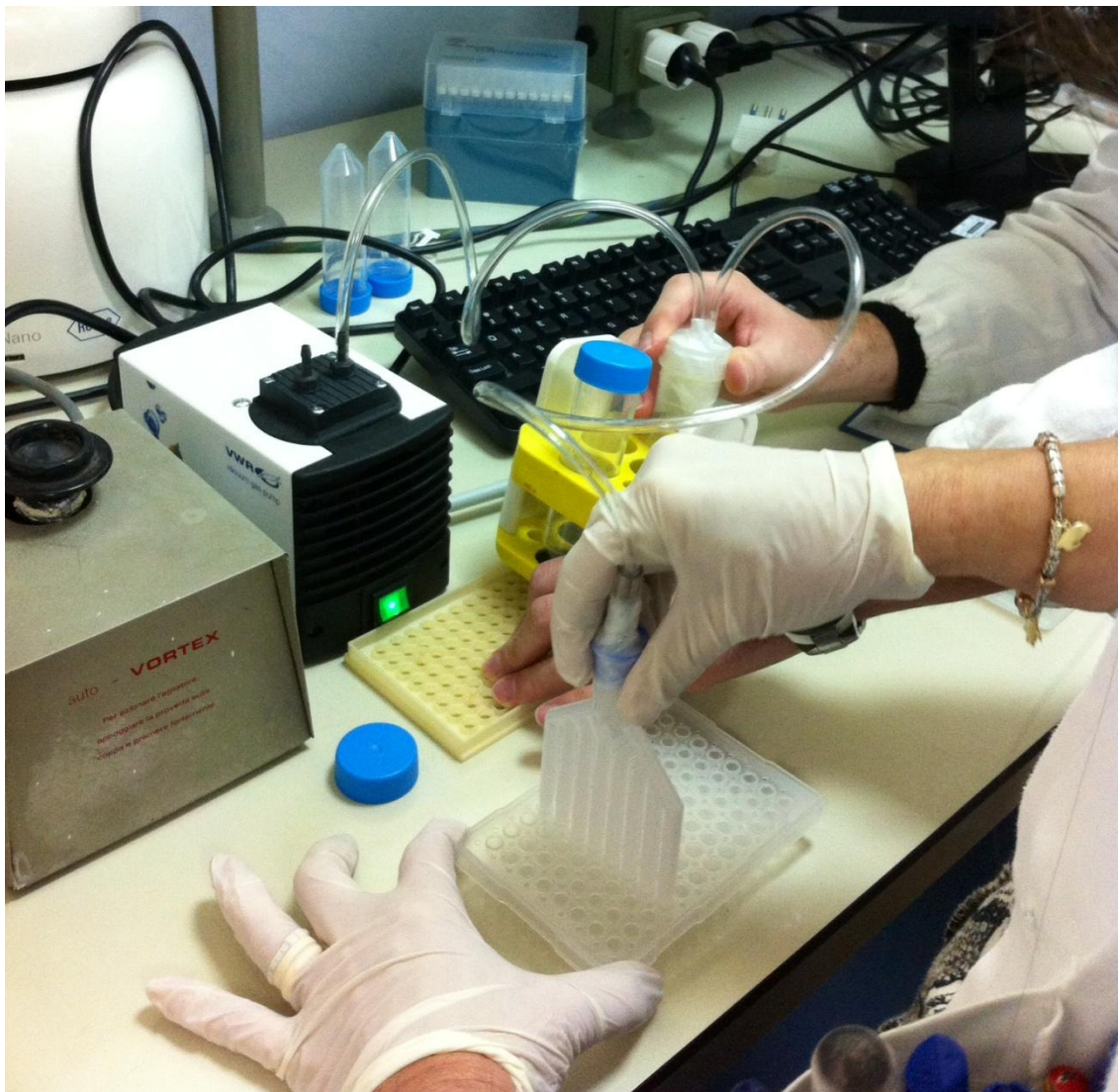


After PCR



Capture Beads

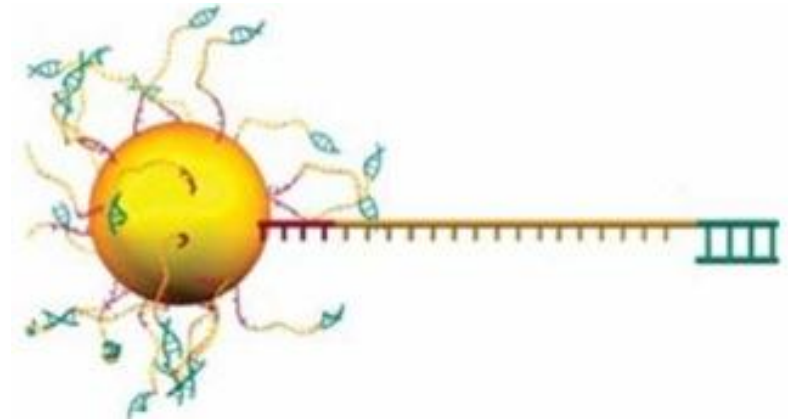
Recupero



**Lavaggi per eliminare
l'emulsione**

Arricchimento

Denaturazione dei ds di
DNA amplificati



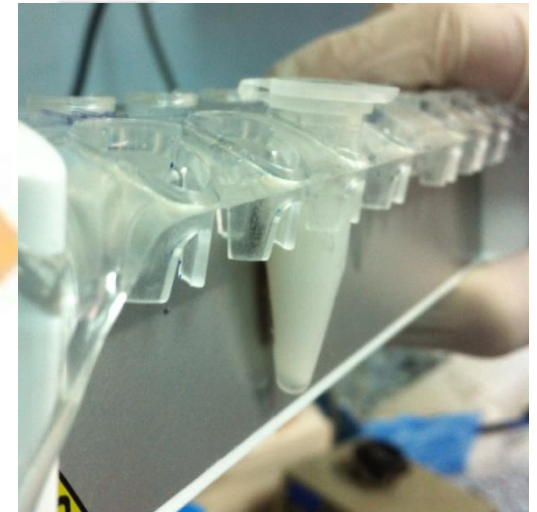
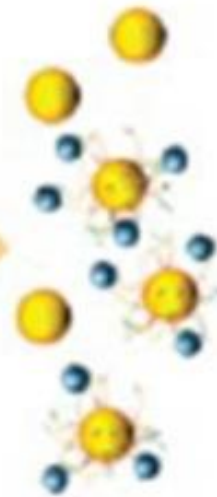
Enrichment
Bead



Bead without
Amplified DNA



Bead with
Amplified DNA

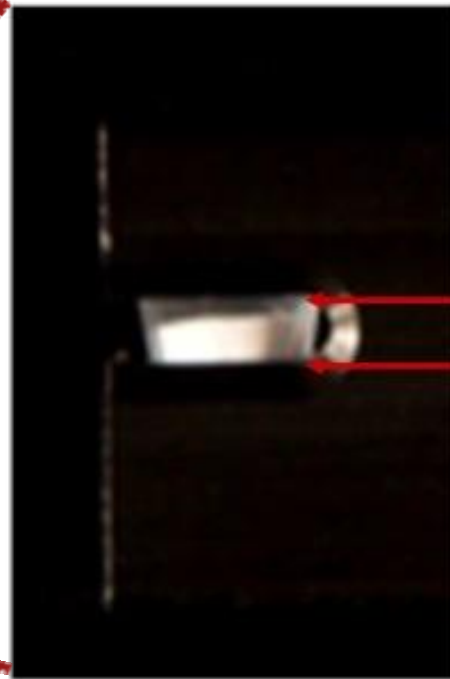


Beads applied to magnet



Annealing dei primer per il sequenziamento

Beads counting



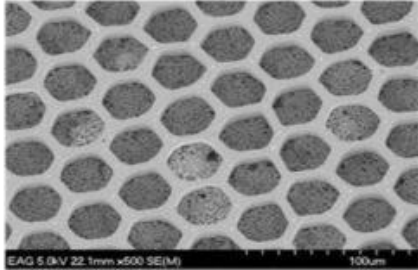
2milioni di beads

20% enrichment

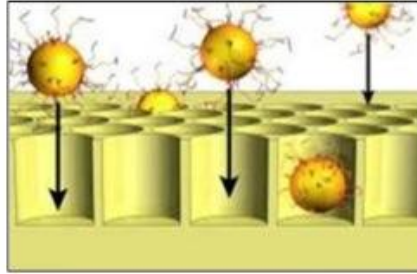
5% enrichment

500 mila beads

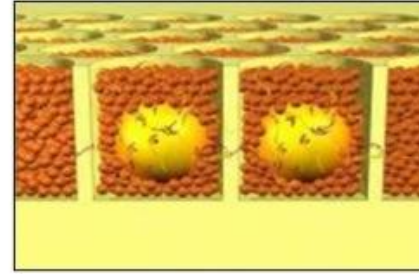
3: Sequenziamento



PicoTiterPlate device



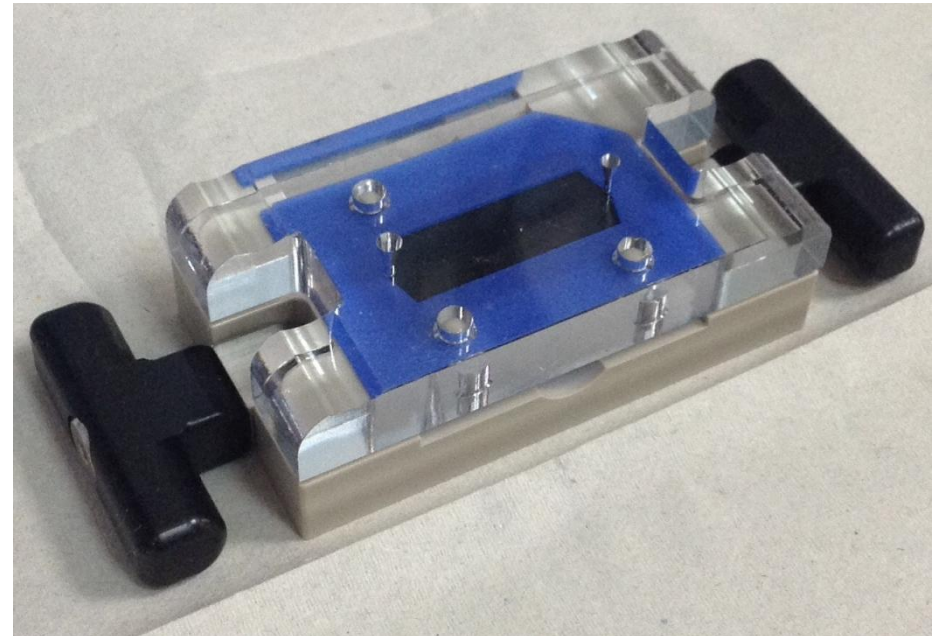
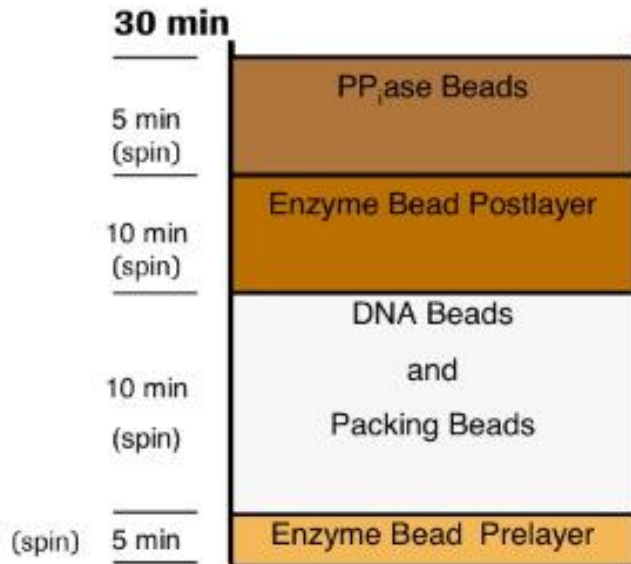
DNA beads are loaded into the wells of the PTP device

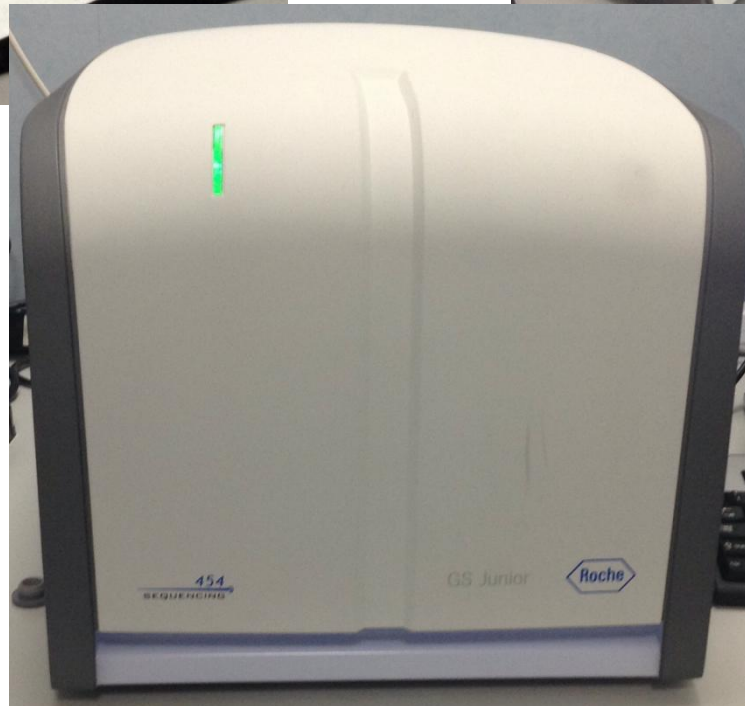
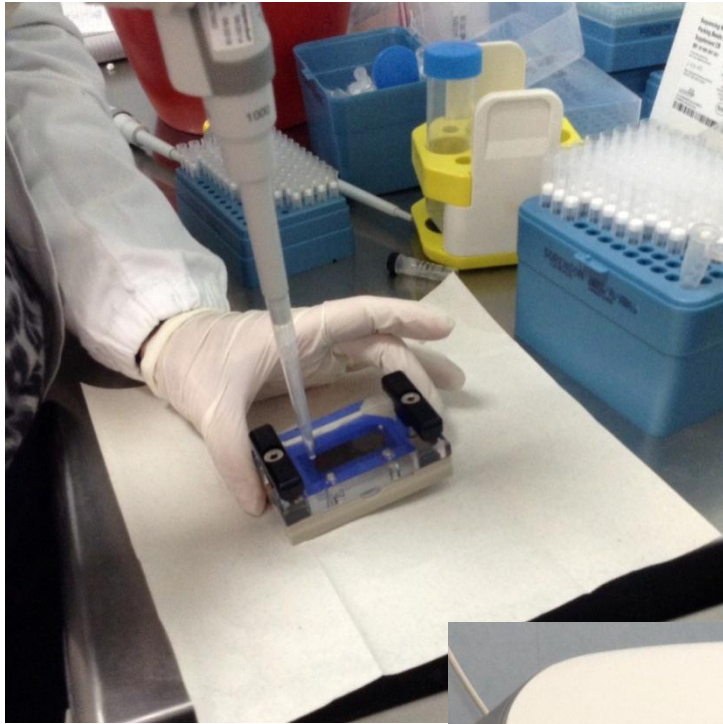


DNA beads packed into wells with surrounding beads and sequencing enzymes.

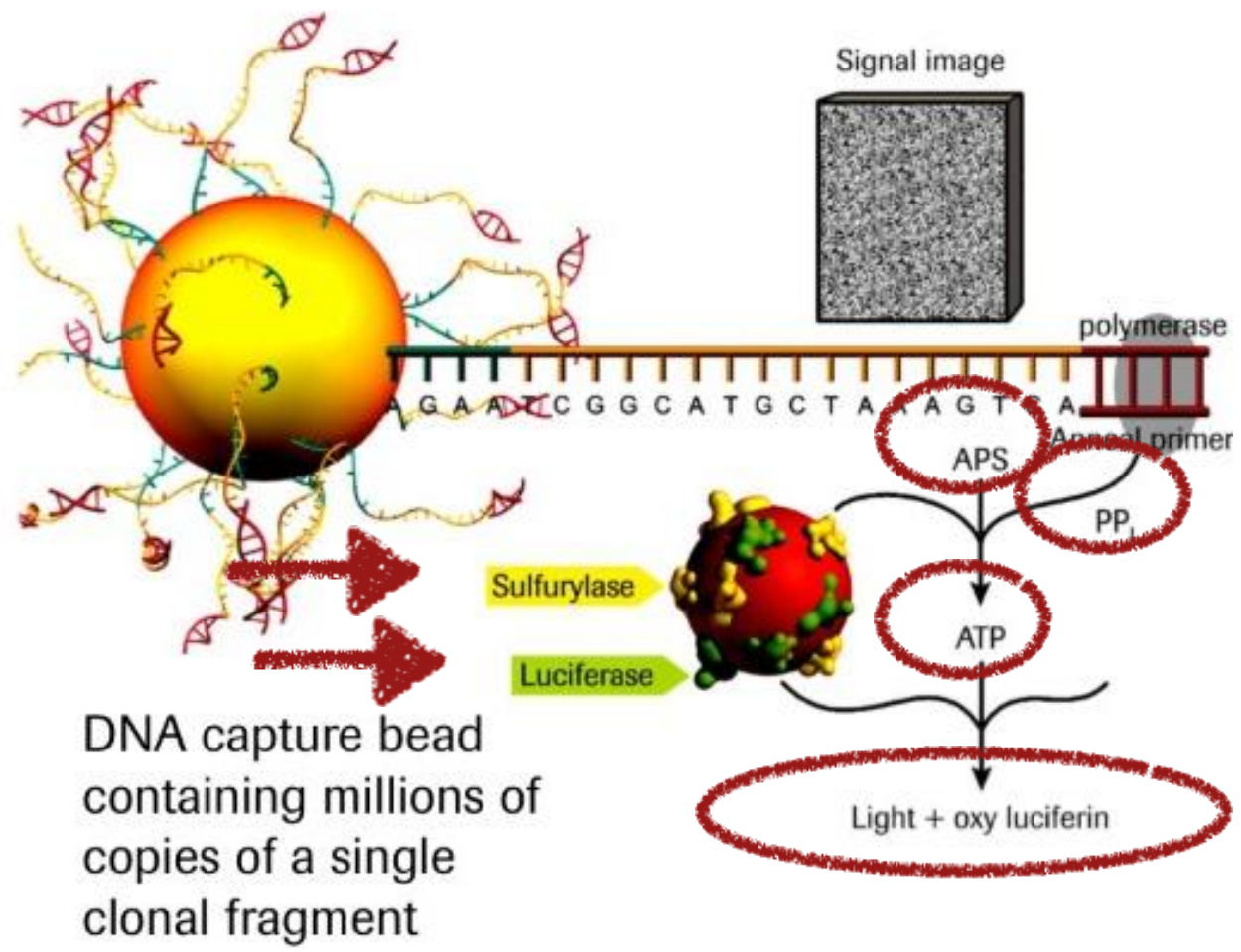
PicoTiter plate

1,5 milioni di pozzetti





L'innovazione del sistema 454-Roche sta nel fatto che con questa tecnica non vengono più utilizzati nucleotidi fluorescenti ma sfrutta il pirofosfato inorganico (PPi), prodotto di scarto della DNA polimerasi. Il pirofosfato inorganico è il substrato dell'enzima sulfurylasi che produce ATP a partire da PPi e APS. A sua volta l'ATP viene prelevato dalla luciferasi che lo utilizza per ossidare la luciferina. Grazie a questa ultima reazione si produce un segnale luminoso. I nucleotidi vengono aggiunti uno per volta proprio perché la fluorescenza è uguale per ciascuno di essi. Il segnale è detectato tramite un fotodetector

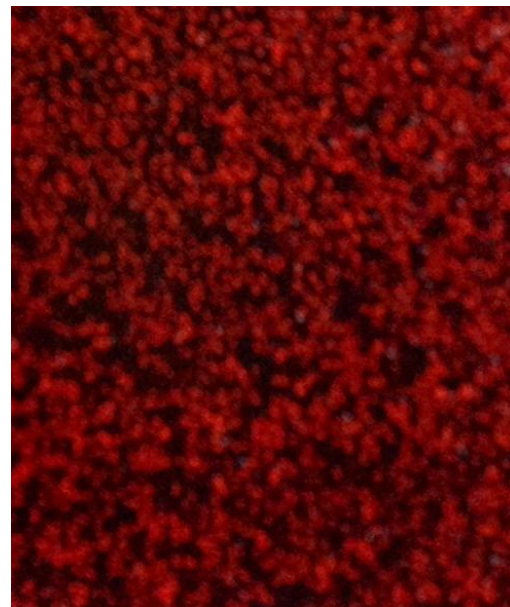




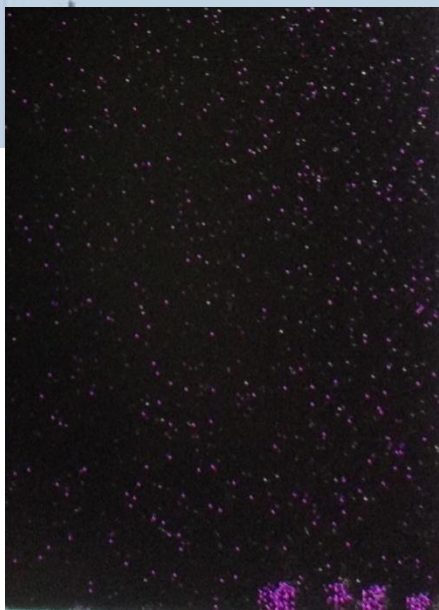
RAW DATA (A/T/C/G)



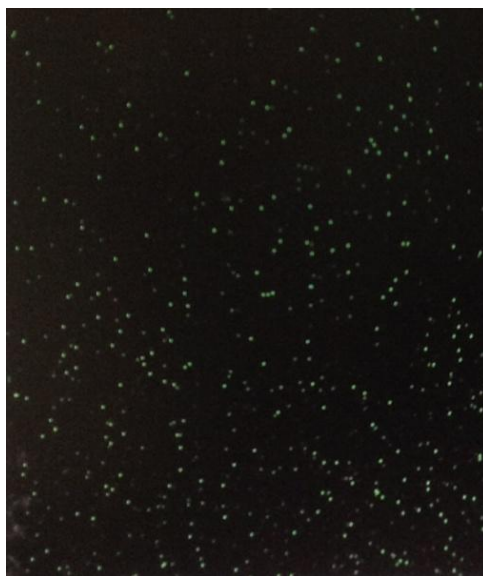
PASSED



FAILED



SEQ NO KEY



CTR PASSED



CTR FAILED

Quali parametri considerare?

Control DNA 400 bp >98%	Raw wells	<20% dot+ mixed	<40% % short quality	% seq passed	n° seq passed
>60- 62%accettabile bene >80%	attorno 200,000-2500000	<6% + <15-20%	< 40%		> 60.000
69,84	237.000	6%	47%	47%	107.250
60,64	258.000	4,20%	54%	41%	101.149
63,68	298.000	8%	53%	38%	109.102
69,43	268.000	13%	58%	28%	70.747
62.66	259.000	27%	50%	22%	53.113
81.88	260.000	8%	40,60%	51,50%	129.206
75,59%	240.000	9%	55%	35%	78.558
68,24%	259.173	6%	37,20%	52,86%	131.651
53,10%	255.402	2%+ 12%	38%	47%	117.848
79,00%	218.181	2%+3%	30%	63,70%	133.167
75,50%	274.347	2%+ 3,5%	43,00%	49,30%	130.594
79.70%	258.624	1%+ 5.3%	46.50%	46.80%	115.570
72,50%	223.898	2,3%+ 3,5%	43,40%	46,80%	98.697
81,10%	233.732	3,7%+ 26,4%	37,70%	30,40%	65.790

ampure	purificazione camp.	primer * PCR di emulsione	n° di molec * capture beads	% di arricchimento camp.
				accettabile 5-20%
1x	1	40 microL	0.8 molecole	10/12 %
0.7x	2	10 microL primer +30 microL H2O	1.4	10/12 %
0.7x	2	40 microL	1.4	12%
1x	2	30 microL primer +10 microL H2O	1.4	15%
0.7x	2	40 microL	1.4	18-20%
0.7x	2	20 microL primer +20 microL H2O	1.4	15%
0.7x	2	20 microL primer +20 microL H2O	1.4	8%
0.7x	2	20 microL primer +20 microL H2O	0,8	5%
0.7x	2	20 microL primer +20 microL H2O	0,8	14%

70.000- 100.000 SEQUENZE

AMPLICONI



SENSIBILITA'

10 ampliconi



100 ampliconi



1000 ampliconi



**10000 Seq X amplicone
(0.1% di sensibilità)**

**1000 Seq X amplicone
(1% di sensibilità)**

**100 Seq X amplicone
(10% di sensibilità)**

HLA Typing Publication from The

Erlich *et al.* *BMC Genomics* 2011, **12**:42
<http://www.biomedcentral.com/1471-2164/12/42>



METHODOLOGY ARTICLE

Open Access

Next-generation sequencing for HLA typing of class I loci

Rachel L Erlich^{1†}, Xiaoming Jia^{1,2†}, Scott Anderson¹, Eric Banks¹, Xiaojiang Gao^{3,4}, Mary Carrington^{3,4}, Namrata Gupta¹, Mark A DePristo¹, Matthew R Henn¹, Niall J Lennon¹, Paul IW de Bakker^{1,5,6,7*}

Abstract

Background: Comprehensive sequence characterization across the MHC is important for successful organ transplantation and genetic association studies. To this end, we have developed an automated sample preparation, molecular barcoding and multiplexing protocol for the amplification and sequence-determination of class I HLA loci. We have coupled this process to a novel HLA calling algorithm to determine the most likely pair of alleles at each locus.

Results: We have benchmarked our protocol with 270 HapMap individuals from four worldwide populations with 96.4% accuracy at 4-digit resolution. A variation of this initial protocol, more suitable for large sample sizes, in which molecular barcodes are added during PCR rather than library construction, was tested on 95 HapMap individuals with 98.6% accuracy at 4-digit resolution.

Conclusions: Next-generation sequencing on the 454 FLX Titanium platform is a reliable, efficient, and scalable technology for HLA typing.

Other HLA Genotyping Publications

- [A novel single cDNA amplicon pyrosequencing method for high-throughput, cost-effective sequence-based HLA class I genotyping.](#) Lank SM, Wiseman RW, Dudley DM, O'Connor DH. (2010) Human Immunology 71(10): 1011-7.
- [Next-Generation Sequencing: The Solution for High-Resolution, Unambiguous HLA Typing.](#) Lind C, Ferriola D, Mackiewicz K, Heron S, Rogers M, Slavich L, Walker R, Hsiao T, McLaughlin L, D'Arcy M, Gai X, Goodridge D, Sayer D, Monos D. (2010) Human Immunology 71(10): 1033-42.
- [High-resolution, high-throughput HLA genotyping by next-generation sequencing.](#) Bentley G, Higuchi R, Hoglund B, Goodridge D, Sayer D, Trachtenberg EA, Erlich HA. (2009) Tissue Antigens 74(5): 393-403.
- [Rapid High-throughput HLA Typing by Massively-parallel Pyrosequencing for High-Resolution Allele Identification.](#) Gabriel C, Danzer M, Hackl C, Kopal G, Hufnagl P, Hofer K, Polin H, Stabentheiner S, Pröll J. (2009) Human Immunology 70(11): 960-4.

GS GType HLA Primer Sets

HLA Typing with the GS FLX and GS Junior Systems

- Two Primer Sets
 - Plate 1 for Medium Resolution
 - Plate 2 for High Resolution (when used with Plate 1)
- Multiplex up to 10 samples per primer plate
- Designed for HLA genotyping on the GS FLX and GS Junior Systems
- Output compatible with 454 System-specific Conexio software – leading provider of HLA Sanger sequencing software
- For life science research use only. Not for use in diagnostic procedures.



GS GType HLA MR Primer Set (blue label)
contains four 96-well plates

GS GType HLA HR Primer Set (yellow label)
contains four 96-well plates

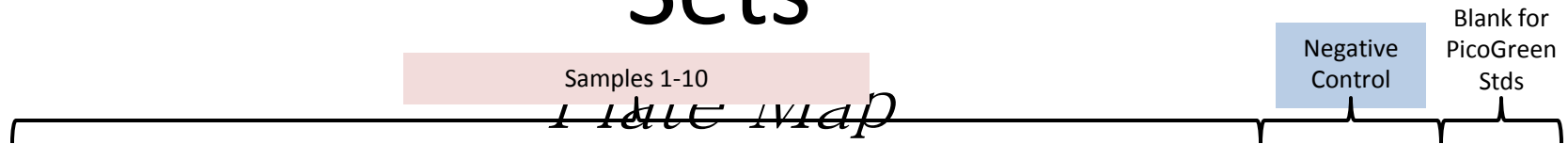
GS GType HLA Primer Sets

Target Exons

- Exons from the HLA Class I and Class II loci targeted in the GS GType MR and HR Primer Sets
- MR plates are used to sequence the exons in column one, or can be used in combination with HR plates to sequence all exons listed

<i>High Resolution (MR + HR)</i>			
<i>Medium Resolution (MR)</i>			
GS GType MR Primer Set		GS GType HR Primer Set	
HLA-A	exons 2, 3	HLA-A	exon 4
HLA-B	exons 2, 3	HLA-B	exon 4
HLA-C	exons 2, 3	HLA-C	exon 4
DQB1	exon 2	DQB1	exon 3
DRB1,3,4,5	exon 2	DPB1	exon 2
		DQA1	exon 2

GS GType HLA MR & HR Primer Sets



Medium Resolution Plate

	MID1	MID3	MID4	MID6	MID7	MID8	MID9	MID10	MID13	MID16	MID11	Blank
	1	2	3	4	5	6	7	8	9	10	11	12
A	HLA-A2	HLA-A2	HLA-A2	HLA-A2	HLA-A2	HLA-A2	HLA-A2	HLA-A2	HLA-A2	HLA-A2	HLA-A2	
B	HLA-A3	HLA-A3	HLA-A3	HLA-A3	HLA-A3	HLA-A3	HLA-A3	HLA-A3	HLA-A3	HLA-A3	HLA-A3	
C	HLA-B2	HLA-B2	HLA-B2	HLA-B2	HLA-B2	HLA-B2	HLA-B2	HLA-B2	HLA-B2	HLA-B2	HLA-B2	
D	HLA-B3	HLA-B3	HLA-B3	HLA-B3	HLA-B3	HLA-B3	HLA-B3	HLA-B3	HLA-B3	HLA-B3	HLA-B3	
E	HLA-C2	HLA-C2	HLA-C2	HLA-C2	HLA-C2	HLA-C2	HLA-C2	HLA-C2	HLA-C2	HLA-C2	HLA-C2	
F	HLA-C3	HLA-C3	HLA-C3	HLA-C3	HLA-C3	HLA-C3	HLA-C3	HLA-C3	HLA-C3	HLA-C3	HLA-C3	
G	DQB1-2	DQB1-2	DQB1-2	DQB1-2	DQB1-2	DQB1-2	DQB1-2	DQB1-2	DQB1-2	DQB1-2	DQB1-2	
H	DRBx-2	DRBx-2	DRBx-2	DRBx-2	DRBx-2	DRBx-2	DRBx-2	DRBx-2	DRBx-2	DRBx-2	DRBx-2	

High Resolution Plate

	MID1	MID3	MID4	MID6	MID7	MID8	MID9	MID10	MID13	MID16	MID11	Blank
	1	2	3	4	5	6	7	8	9	10	11	12
A	HLA-A4	HLA-A4	HLA-A4	HLA-A4	HLA-A4	HLA-A4	HLA-A4	HLA-A4	HLA-A4	HLA-A4	HLA-A4	
B	HLA-B4	HLA-B4	HLA-B4	HLA-B4	HLA-B4	HLA-B4	HLA-B4	HLA-B4	HLA-B4	HLA-B4	HLA-B4	
C	HLA-C4	HLA-C4	HLA-C4	HLA-C4	HLA-C4	HLA-C4	HLA-C4	HLA-C4	HLA-C4	HLA-C4	HLA-C4	
D	DPB1-2	DPB1-2	DPB1-2	DPB1-2	DPB1-2	DPB1-2	DPB1-2	DPB1-2	DPB1-2	DPB1-2	DPB1-2	
E	DQA1-2	DQA1-2	DQA1-2	DQA1-2	DQA1-2	DQA1-2	DQA1-2	DQA1-2	DQA1-2	DQA1-2	DQA1-2	
F	DQB1-3	DQB1-3	DQB1-3	DQB1-3	DQB1-3	DQB1-3	DQB1-3	DQB1-3	DQB1-3	DQB1-3	DQB1-3	
G												
H												

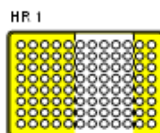
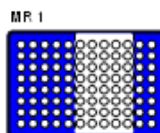
Options for Sequencing *GS Junior System*



**Number of
Samples**

HR on GS Junior

$\frac{1}{2}$ MR plate + $\frac{1}{2}$ HR plate =
5 samples @ 14 Amplicons / sample

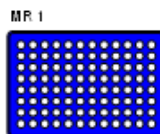


Single size gasket
5 HR samples
(or 10 HR samples
on two Runs)

5

MR on GS Junior

1 MR plate =
10 samples @ 8 Amplicons / sample



Single size gasket
10 MR samples

10

ATF 454 Analysis Software from Conexio

Automatic Allele Assignment



The screenshot shows the ATF 454 Analysis Software interface. The main window displays a sequence alignment and a table of mismatches (MM0, MM2, MM3) for HLA-B. The table shows that for the B*15:10:01/B*38:01:01 assignment, there are zero mismatches (MM0, MM2, MM3), indicating an unambiguous genotype assignment.

Sample	Sequence	Allele 1	Allele 2	MM 0	MM 2	MM 3	Difference
Sample 01	CCTCCTCCGCGGGCATACCAAGTTCGCTACGACGGC	B*15:10:01	B*38:01:01	0	0	0	
Sample 02	CCTCCTCCGCGGGCATACCAAGTTCGCTACGACGGC	B*15:10:01	B*38:01:02	1	NC	NC	
Sample 03	CCTCCTCCGCGGGCATACCAAGTTCGCTACGACGGC	B*15:10:01	B*38:01:03	1	NC	NC	
Sample 04	CCTCCTCCGCGGGCATACCAAGTTCGCTACGACGGC	B*15:10:01	B*38:01:04	1	NC	NC	
Sample 05	CCTCCTCCGCGGGCATACCAAGTTCGCTACGACGGC	B*15:10:01	B*38:05	1	NC	NC	
Sample 06	CCTCCTCCGCGGGCATACCAAGTTCGCTACGACGGC	B*15:10:01	B*38:09	1	NC	NC	
Sample 07* B	CCTCCTCCGCGGGCATACCAAGTTCGCTACGACGGC	B*15:10:01	B*38:12	1	NC	NC	
Sample 08	CCTCCTCCGCGGGCATACCAAGTTCGCTACGACGGC	B*15:10:01	B*38:13	1	NC	NC	
Sample 09	CCTCCTCCGCGGGCATACCAAGTTCGCTACGACGGC	B*15:10:01	B*38:14	1	NC	NC	
Sample 10	CCTCCTCCGCGGGCATACCAAGTTCGCTACGACGGC	B*15:10:01	B*38:19	1	NC	NC	
		B*15:10:01	B*38:21	1	NC	NC	
		B*15:10:02	B*38:01:01	1	NC	NC	
		B*15:18:01	B*38:01:01	1	NC	NC	
		B*15:23	B*39:05:01	1	NC	NC	
		B*15:10:01	B*38:10	2	NC	NC	
		B*15:10:01	B*38:11	2	NC	NC	
		B*15:10:01	B*38:15	2	NC	NC	
		B*15:10:01	B*38:16	2	NC	NC	
		B*15:10:01	B*38:20	2	NC	NC	

Zero mismatches
= Unambiguous
genotype
assignment

Example shown: For HLA-B, the Conexio ATF interface shows an unambiguous genotype assignment (zero mismatches in columns MM0, MM2, MM3) for B*15:10:01/B*38:01:01.

5 samples 20 ampliconi
Sensibilità <1%

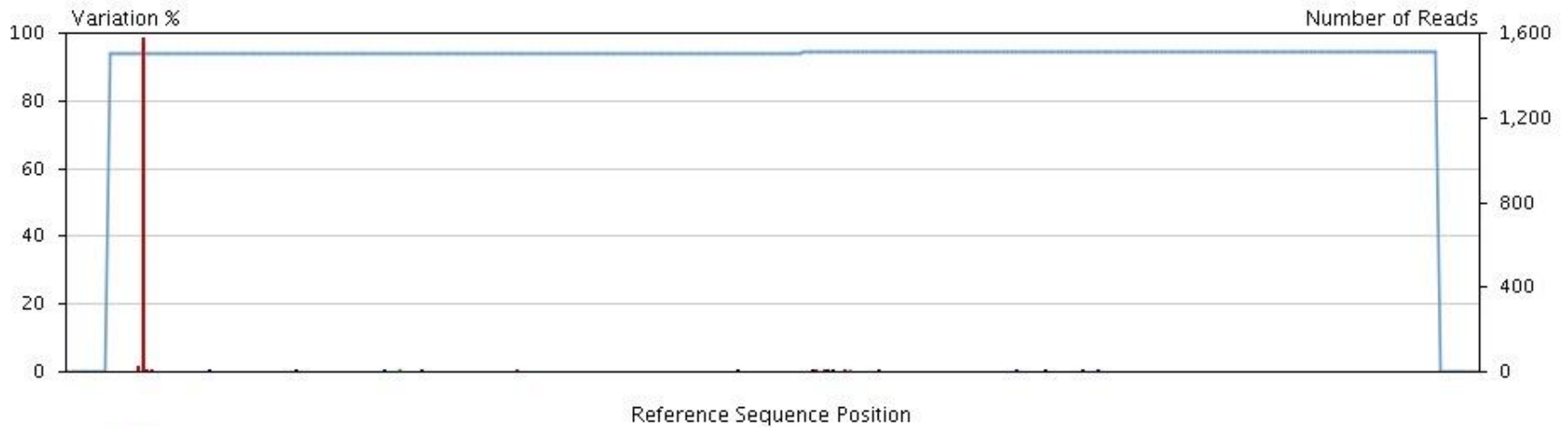
12 samples 84 ampliconi
Sensibilità 1%

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8	Sample 9	Sample 10	Sample 11	Sample 12
	MID1	MID2	MID3	MID4	MID19	MID6	MID8	MID10	MID13	MID18	MID15	MID16
A	Exon 4	Exon 4	Exon 4	Exon 4	Exon 4	Exon 4	Exon 4	Exon 4	Exon 4	Exon 4	Exon 4	Exon 4
B	Exon 5	Exon 5	Exon 5	Exon 5	Exon 5	Exon 5	Exon 5	Exon 5	Exon 5	Exon 5	Exon 5	Exon 5
C	Exon 6	Exon 6	Exon 6	Exon 6	Exon 6	Exon 6	Exon 6	Exon 6	Exon 6	Exon 6	Exon 6	Exon 6
D	Exon 7	Exon 7	Exon 7	Exon 7	Exon 7	Exon 7	Exon 7	Exon 7	Exon 7	Exon 7	Exon 7	Exon 7
E	Exon 3	Exon 3	Exon 3	Exon 3	Exon 3	Exon 3	Exon 3	Exon 3	Exon 3	Exon 3	Exon 3	Exon 3
F	Ex 8.1	Ex 8.1	Ex 8.1	Ex 8.1	Ex 8.1	Ex 8.1	Ex 8.1	Ex 8.1	Ex 8.1	Ex 8.1	Ex 8.1	Ex 8.1
G	Ex 8.2	Ex 8.2	Ex 8.2	Ex 8.2	Ex 8.2	Ex 8.2	Ex 8.2	Ex 8.2	Ex 8.2	Ex 8.2	Ex 8.2	Ex 8.2
H											MID 17 Exon 7 NC*	MID 17 Ex 8.2 NC*

11 samples 88 ampliconi
Sensibilità 1%

[illegible]

Software AVA



MRD Detection in ALL Patients

Detection and selection of clonal Ig/TCR gene rearrangements at diagnosis:

Ig/TCR PCR

Sequencing of clonal rearrangements

Sequence interpretation

Selection of MRD-PCR targets

Guidelines BIOMED



RQ-PCR sensitivity testing:

Design of allele-specific oligonucleotide primers

RQ-PCR analysis of dilution series of diagnostic sample

RQ-PCR data interpretation: quantitative range and sensitivity

Guidelines ESG-MRD-ALL



MRD analysis of follow-up samples

Control gene RQ-PCR analysis

MRD-PCR target RQ-PCR analysis

RQ-PCR MRD data interpretation

Guidelines ESG-MRD-ALL

1. DNA preparation

(2-3 days)

- a. BM sampling at diagnosis (≥ 5 ml)
- b. MNC-density gradient separation (1×10^7 cells)
- c. Genomic DNA extraction ($\geq 10 \mu\text{g}$)

2. MRD PCR target identification

(1-2 weeks)

- a. PCR-heteroduplex analysis
- b. Sequencing of clonal rearrangements
- c. Sequence interpretation
- d. Selection of MRD-PCR targets

3. RQ-PCR design and sensitivity testing

(1-2 weeks)

- a. Design of allele-specific oligonucleotide primers
- b. RQ-PCR analysis of dilution series of diagnostic sample
- c. RQ-PCR data interpretation

4. MRD analysis of follow-up samples

(1-2 weeks)

- a. RQ-PCR analysis of follow-up samples (control gene)
- b. RQ-PCR analysis of follow-up samples (Ig/TCR targets)
- c. RQ-PCR data interpretation
- d. Calculation of MRD level

Approccio “standard”

BIOMED 2-RQPCR

Screening

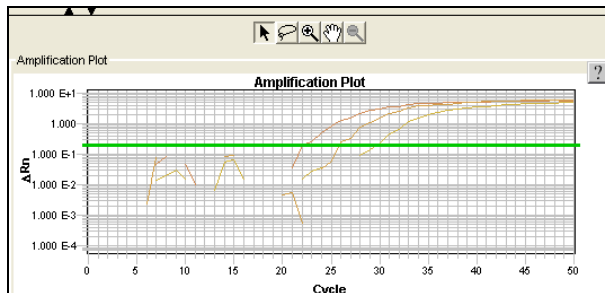
IGH
DH
IGK-Kde



Selezione di 2 marcatori

MRD

Oligo paziente specifico
Amplificazione specifica
con metodologia TaqMan



Approccio “NGS”

Screening

IGH
DH
IGK-Kde



MRD

IGH
DH
IGK-Kde



CD-HIT

Representative Sequences.

[Home](#)
[Download](#)
[Web Servers](#)
[Doc. & References](#)
[Applications](#)
[Community development](#)
[Contact Us](#)

CD-HIT is a very widely used program for clustering and comparing protein or nucleotide sequences. CD-HIT was originally developed by Dr. Weizhong Li at Dr. Adam Godzik's Lab at the Burnham Institute (now Sanford-Burnham Medical Research Institute).

CD-HIT is very fast and can handle extremely large databases. CD-HIT helps to significantly reduce the computational and manual efforts in many sequence analysis tasks and aids in understanding the data structure and correct the bias within a dataset.

The CD-HIT package has CD-HIT, CD-HIT-2D, CD-HIT-EST, CD-HIT-EST-2D, CD-HIT-454, CD-HIT-PARA, PSI-CD-HIT, CD-HIT-OTU, CD-HIT-LAP, CD-HIT-DUP and over a dozen scripts.

- CD-HIT (CD-HIT-EST) clusters similar proteins (DNAs) into clusters that meet a user-defined similarity threshold.
- CD-HIT-2D (CD-HIT-EST-2D) compares 2 datasets and identifies the sequences in db2 that are similar to db1 above a threshold.
- CD-HIT-454 identifies natural and artificial duplicates from pyrosequencing reads.
- CD-HIT-OTU clusters rRNA tags into OTUs.
- CD-HIT-DUP identifies duplicates from single or paired Illumina reads.
- CD-HIT-LAP identifies overlapping reads.

The usage of other programs and scripts can be found in CD-HIT user's guide.

CD-HIT is currently maintained by the Dr. Li's group (<http://weizhongliab.ucsd.edu/>). We thank the support from National Center for Research Resources (Grant # 1R01RR020530, 2008-2011). We thank all users that report bugs, give us suggestions and comments.

NEWS

(June 2013) We were awarded an "AISS in Education Research Grant Award" by Amazon. This will support us to develop cloud-based CD-HIT applications.

(May 2013) We invite users to help test [cd-hit-2d](#) before public release.

(October 2012) The new CD-HIT paper was just published at [Bioinformatics](#).

(July 2012) A paper was just published at [Bioinformatics](#). This paper describes several clustering applications in metagenomic data analysis.

(June 2011) [cd-hit-2d](#) is a special cd-hit extension for clustering rRNA tags into OTUs. It is very fast and very accurate.

(July 2010) [cd-hit](#) [Google Code](#) is a new Google Code project created for releasing the latest development version of CD-HIT. Usually new minor versions will be released as soon as bug fixings or improvements become available.

(Oct. 2009) CD-HIT-454 is a new program to identify exact duplicates and near identical duplicates in pyrosequencing reads. [CD-HIT-454 \(latest\)](#), [CD-HIT-454 \(downloads\)](#).

(September 2009) CD-HIT [web server](#) is now available to run cd-hit or download pre-pre-calculated clusters.

(December 2008) I made some major updates including several very useful new options for clustering such as alignment coverage control, search between local and global sequence identity. Please check the newest release and have a try. (Weizhong Li)

19 cluster

Cluster 0, No. sequences: 17822
 Representative: I6JKX5O01CVONS
 >I6JKX5O01CVONS length=545
 xy=1063_3590

CD-HIT Suite: Biological Sequence Clustering and Comparison

[Server home](#)
[cd-hit](#)
[cd-hit-est](#)
[h-cd-hit](#)
[h-cd-hit-est](#)
[cd-hit-2d](#)
[cd-hit-est-2d](#)
[result](#)
[calculated clusters](#)

Sequence file and databases
Load Query Fasta file from your computer: Nessun file selezionato
☐ Incorporate annotation info at header line

Sequence Identity Parameters
Sequence identity cut-off:

Algorithm Parameters
-r: comparing both strands ☐ No ☐ Yes
-G: use global sequence identity ☐ No ☐ Yes
-g: sequence is clustered to the best cluster that meet the threshold ☐ No ☐ Yes
-b: bandwidth of alignment

Alignment Coverage Parameters
-aL: minimal alignment coverage (fraction) for the longer sequence
-AL: maximum unaligned part (amino acids/bases) for the longer sequence
-aS: minimal alignment coverage (fraction) for the shorter sequence
-AS: maximum unaligned part (amino acids/bases) for the shorter sequence
-s: minimal length similarity (fraction)
-S: maximum length difference in amino acids/bases(-S)

Mail address for job checking
Give your mail address:

VH4-55*02	DH3-09*01	JH6*03
VH4-55*02	DH3-09*01	JH6*03
IGKV1-37*01		
IGKV1-37*01-Kde		
IGKV4-1*01		
IGKV4-1*01-Kde		
IGKV2-30*01		
IGKV2-30*01-Kde		
Vk2-30-Kde		
Vk2-30-Kde		
Vk3D-20		
Vk3-20-Kde		
DH2-2*01-JH5*01		
VH6-1*02	DH2-2*02	JH5*02
VH3-11*06	DH5-18*01	JH4*02
VH3-11*06	DH5-18*01	JH4*02
VH2-5*02	DH3-22*01	JH4*02
VH2-5*02	DH3-22*01	JH4*02
IGKV1D-33*01		
IGKV1D-33*01-kde		

Huang Y at al.
 Bioinformatics, 2010(26): 680-682

Cd-hit-est-2d

11/ 13.000

EuroMRD-
QC: 5×10^{-4}

Screening

VH6

Sequence file and databases

CD-HIT-EST-2D compares 2 nucleotide datasets (db1, db2). It identifies the sequences in db2 that are similar to db1 at a certain threshold.

Choose db1

Load search database (in Fasta format): Nessun file selezionato

Choose db2

Load Query Fasta file from your computer: Nessun file selezionato

Sequence Identity Parameters

Sequence identity cut-off

0.9

Algorithm Parameters

-r: comparing both strands

☒ No ☐ Yes

-G: use global sequence identity

☐ No ☒ Yes

-g: sequence is clustered to the best cluster that meet the threshold

☐ No ☒ Yes

-b: bandwidth of alignment

20

Alignment Coverage Parameters

-aL: minimal alignment coverage (fraction) for the longer sequence

0.0

-AL: maximum unaligned part (amino acids/bases) for the longer sequence

unlimited

-aS: minimal alignment coverage (fraction) for the shorter sequence

0.0

-AS: maximum unaligned part (amino acids/bases) for the shorter sequence

unlimited

-s: minimal length similarity (fraction)

0.0

-S: maximum length difference in amino acids/bases (-S)

unlimited

Length Control Parameters

Length difference cutoff (fraction)

1.0

Length difference cutoff (amino acids/bases)

0

Mail address for job checking

Give your mail address:

0.98

```
1. JCKBR3K01AJN8J, length: 331, identity: +/98% VH6-1+01 DH7-27+01 JH5+02
2. JCKBR3K01CCQCL, length: 331, identity: +/99% VH6-1+01 DH7-27+01 JH5+02
3. JCKBR3K01COK03, length: 331, identity: +/99% VH6-1+01 DH7-27+01 JH5+02
4. JCKBR3K01C30TM, length: 331, identity: +/99% VH6-1+01 DH7-27+01 JH5+02
5. JCKBR3K01C2FEG, length: 331, identity: +/99% VH6-1+01 DH7-27+01 JH5+02
6. JCKBR3K01AFLGD, length: 331, identity: +/99% VH6-1+01 DH7-27+01 JH5+02
7. JCKBR3K01CF19P, length: 331, identity: +/99% VH6-1+01 DH7-27+01 JH5+02
8. JCKBR3K01A6NAM, length: 331, identity: +/99% VH6-1+01 DH7-27+01 JH5+02
9. JCKBR3K01CUNYJ, length: 331, identity: +/98% VH6-1+01 DH7-27+01 JH5+02
10 JCKBR3K01DDP2P, length: 330, identity: +/99% VH6-1+01 DH7-27+01 JH5+02
11 JCKBR3K01BBOOL, length: 330, identity: +/99% VH6-1+01 DH7-27+01 JH5+02
```

0.97

```
1. JCKBR3K01AYXLV, length: 338, identity: +/97% VH6-1+01 DH7-27+01 JH5+02
2. JCKBR3K01AJN8J, length: 331, identity: +/98%
3. JCKBR3K01CCQCL, length: 331, identity: +/99%
4. JCKBR3K01COK03, length: 331, identity: +/99%
5. JCKBR3K01C30TM, length: 331, identity: +/99%
6. JCKBR3K01C2FEG, length: 331, identity: +/99%
7. JCKBR3K01AFLGD, length: 331, identity: +/99%
8. JCKBR3K01CF19P, length: 331, identity: +/99%
9. JCKBR3K01A6NAM, length: 331, identity: +/99%
10. JCKBR3K01CUNYJ, length: 331, identity: +/98%
11. JCKBR3K01DDP2P, length: 330, identity: +/99%
12. JCKBR3K01BBOOL, length: 330, identity: +/99%
```

[Leukemia](#). 2014 Jun;28(6):1299-307. doi: 10.1038/leu.2013.375. Epub 2013 Dec 17.

Next-generation sequencing and real-time quantitative PCR for minimal residual disease detection in B-cell disorders.

[Ladetto M](#)¹, [Brüggemann M](#)², [Monitillo L](#)¹, [Ferrero S](#)¹, [Pepin F](#)³, [Drandi D](#)¹, [Barbero D](#)¹, [Palumbo A](#)¹, [Passera R](#)⁴, [Boccadoro M](#)¹, [Ritgen M](#)², [Gökbuget N](#)⁵, [Zheng J](#)³, [Carlton V](#)³, [Trautmann H](#)², [Faham M](#)³, [Pott C](#)².

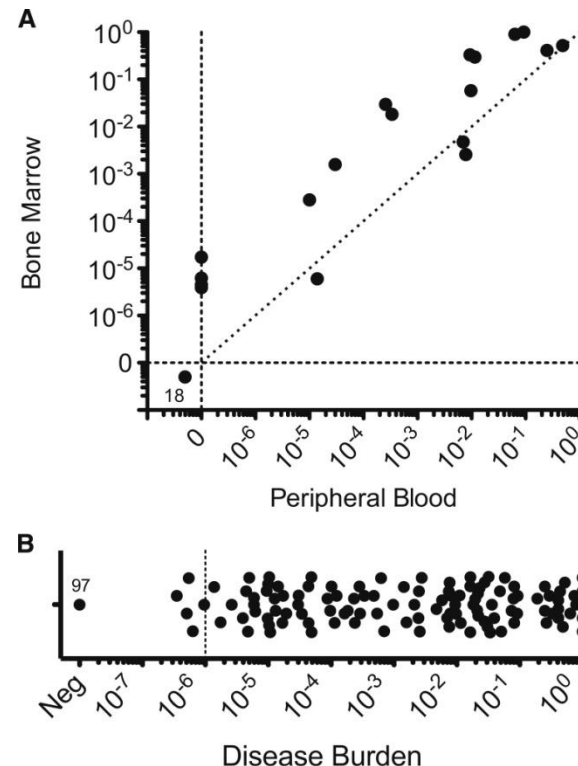
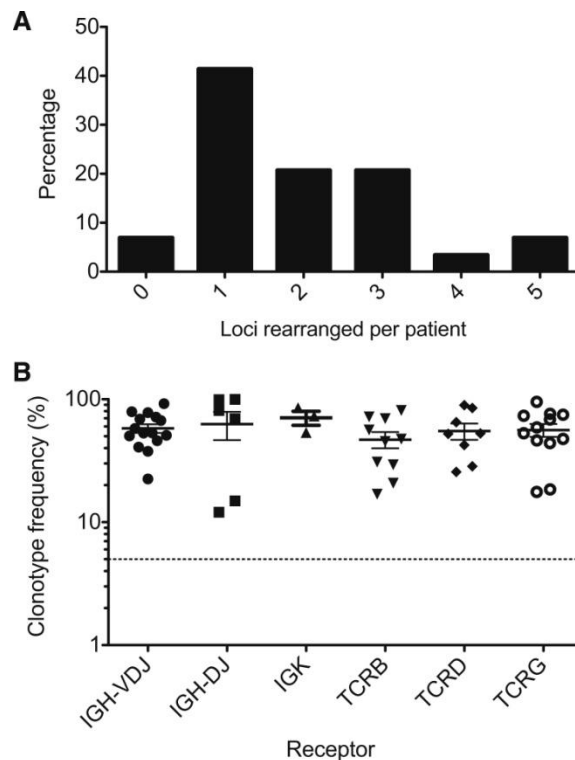
NGS showed at least the same level of sensitivity as allele-specific oligonucleotides-PCR, without the need for patient-specific reagents. We conclude that NGS is an effective tool for MRD monitoring in ALL, MCL and MM. Prospective comparative analysis of unselected cases is required to validate the clinical impact of NGS-based MRD assessment

*Immunoglobulin and T Cell Receptor Gene High-Throughput Sequencing
Quantifies Minimal Residual Disease in Acute Lymphoblastic Leukemia and
Predicts Post-Transplantation Relapse and Survival*

*Aaron C. Logan, Nikita Vashi, Malek Faham, Victoria Carlton, Katherine Kong,
Ismael Buño, Jianbiao Zheng, Martin Moorhead, Mark Klinger, Bing Zhang,
Amna Waqar, James L. Zehnder, David B. Miklos*
Biology of Blood and Marrow Transplant

Volume 20, Issue 9, Pages 1307-1313 (September 2014)

DOI: 10.1016/j.bbmt.2014.04.018



[Blood](#). 2015 May 28;125(22):3501-8. doi: 10.1182/blood-2014-12-615757. Epub 2015 Apr 10.

IgH-V(D)J NGS-MRD measurement pre- and early post-allotransplant defines very low- and very high-risk ALL patients.

[Pulsipher MA](#)¹, [Carlson C](#)², [Langholz B](#)³, [Wall DA](#)⁴, [Schultz KR](#)⁵, [Bunin N](#)⁶, [Kirsch I](#)⁷, [Gastier-Foster JM](#)⁸, [Borowitz M](#)⁹, [Desmarais C](#)⁷, [Williamson D](#)⁷, [Kalos M](#)¹⁰, [Grupp SA](#)¹¹.

- Absence of detectable IgH-V(D)J NGS-MRD pre-HCT defines good-risk patients potentially eligible for less intense treatment approaches. Post-HCT NGS-MRD is highly predictive of relapse and survival, suggesting a role for this technique in defining patients early who would be eligible for post-HCT interventions.

Altre applicazioni...

- **Sequenziamento del trascrittoma-RNAseq**
- **Sequenziamento di porzioni di DNA legate da proteine regolatorie-ChIPseq**
- **Metilazione**