AIBT Winter School
Ravascletto (UD) 3-5 Dicembre 2015

NGS: vantaggi, ambiti e ricadute.

Sandro Orrù

UNIVERSITA' DEGLI STUDI DI CAGLIARI
Dipartimento di Scienze Mediche
Genetica Medica

Ospedale "R. Binaghi" Via Is Guadazzonis 3, 09126 Cagliari

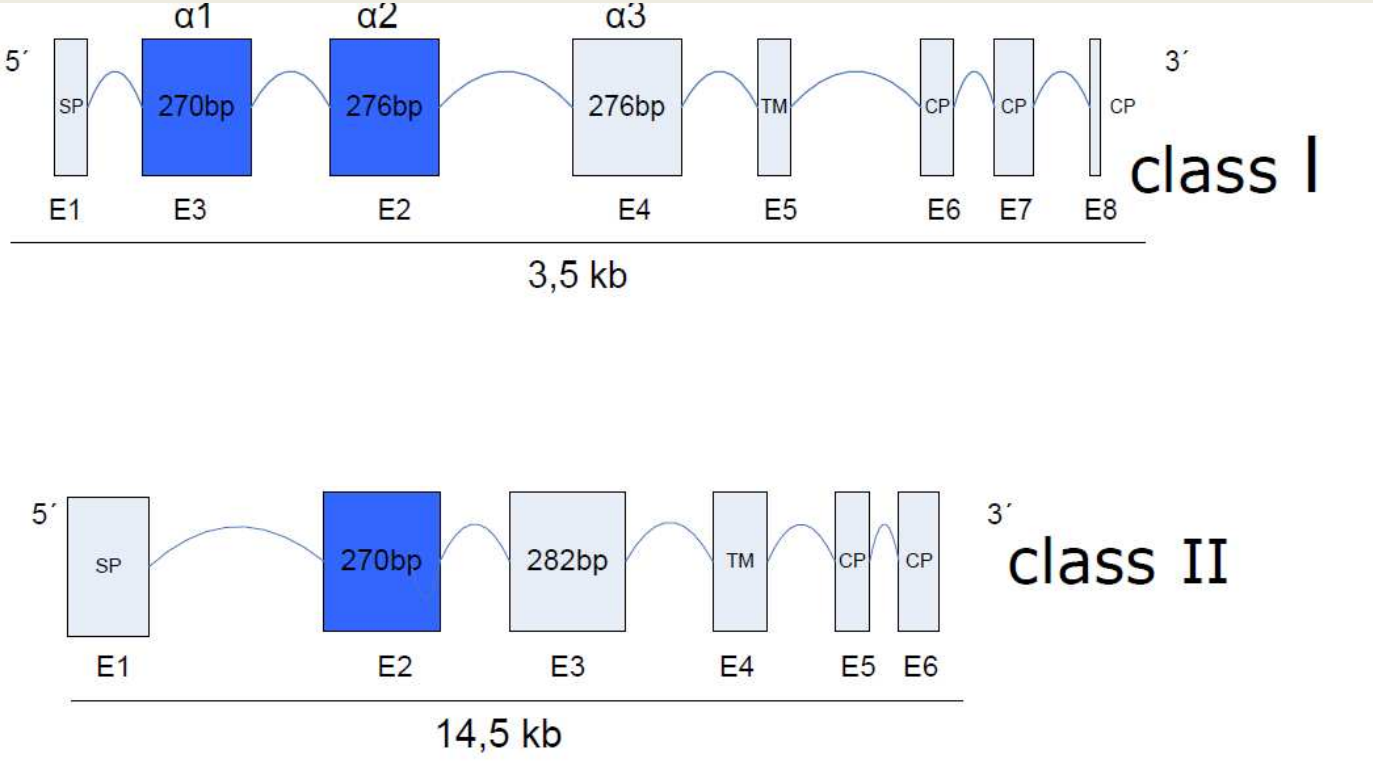
HLA Typing Methods

- Serologic assays
  - Microlymphocytotoxicity test
- Cellular assays
  - Mixed lymphocyte culture
- Molecular assays
  - Sequence-specific primer (SSP)
  - Sequence-specific oligonucleotide probe (SSOP)
  - Sequence-based testing (SBT)


Edit Reports Update View Window Help
Exon 2
Exon 3
Exon 4
Exon 5
A2F_B01
A2R_B02
A3F_B03
A3R_B04
A4F_B05
A4R_B06
RCBGTRGATRGAGCRGRAGRRKSYDGMGTATTRGGAYVNGVAVACRSGGMAHRTRARGDBCCVSTCACA
GCCGTGGATAGAGCAGGAGGGKCCGGAGTATTGGGACSRGGAGACACGGAAWRTGAAGGCCCACTCACA
A: 393 C: 522 G: 599 T: 469
A: 79 C: 160 G: 170 T: 206
Start: 2 (74) Exon 2 1
Stop: 823 (895) Exon 4 276
Allele 1
Allele 2
MM
A*01010101
A*02010101
A*02010102L
A*02010103
A*01010102N
A*0114
A*9201
A*0236
A*3604
A*020103
A*020104
A*020105
A*020106
A*020107
A*020108
A*020109
A*020110
A*020111
A*020112
A*020113
A*020114
A*020115
A*020118
A*020119
A*020121
A*020122
A*0204
A*0207
A*0209
A*022001
A*022002
A*0222
Master
No Offset
82.1
244
BCS
MM
A*01:01, 02:01
NUM

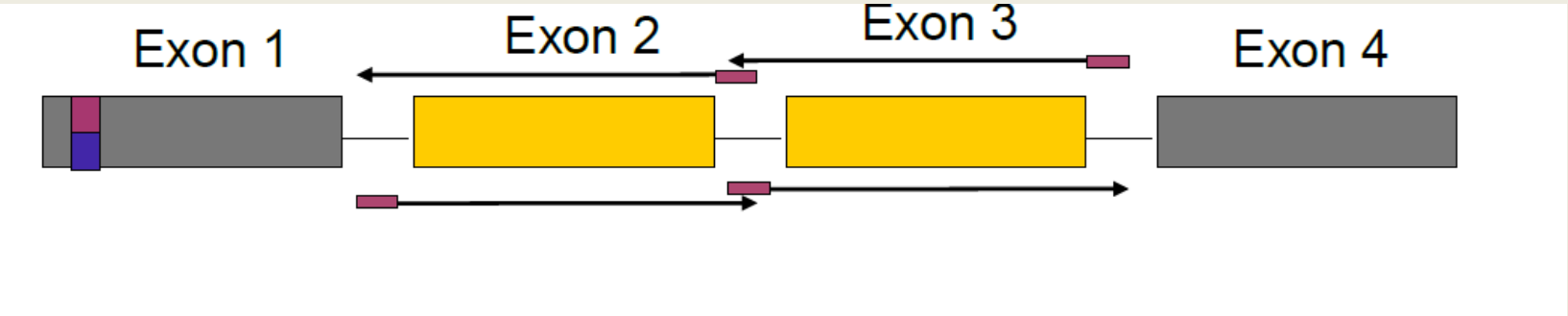
Sequenziamento secondo Sanger dei loci HLA

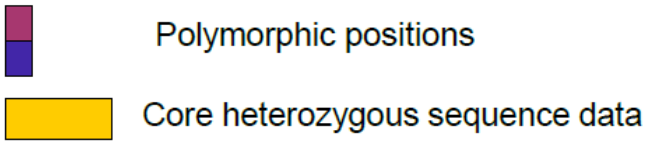
A causa dei costi e della laboriosità del metodo, l' analisi dei loci HLA è solo raramente condotta sull' intera regione codificante.



Allele ambiguity

outlier mutations: allele ambiguitiy
Si verifica quando un polimorfismo che distingue un allele cade fuori dalla regione esaminata





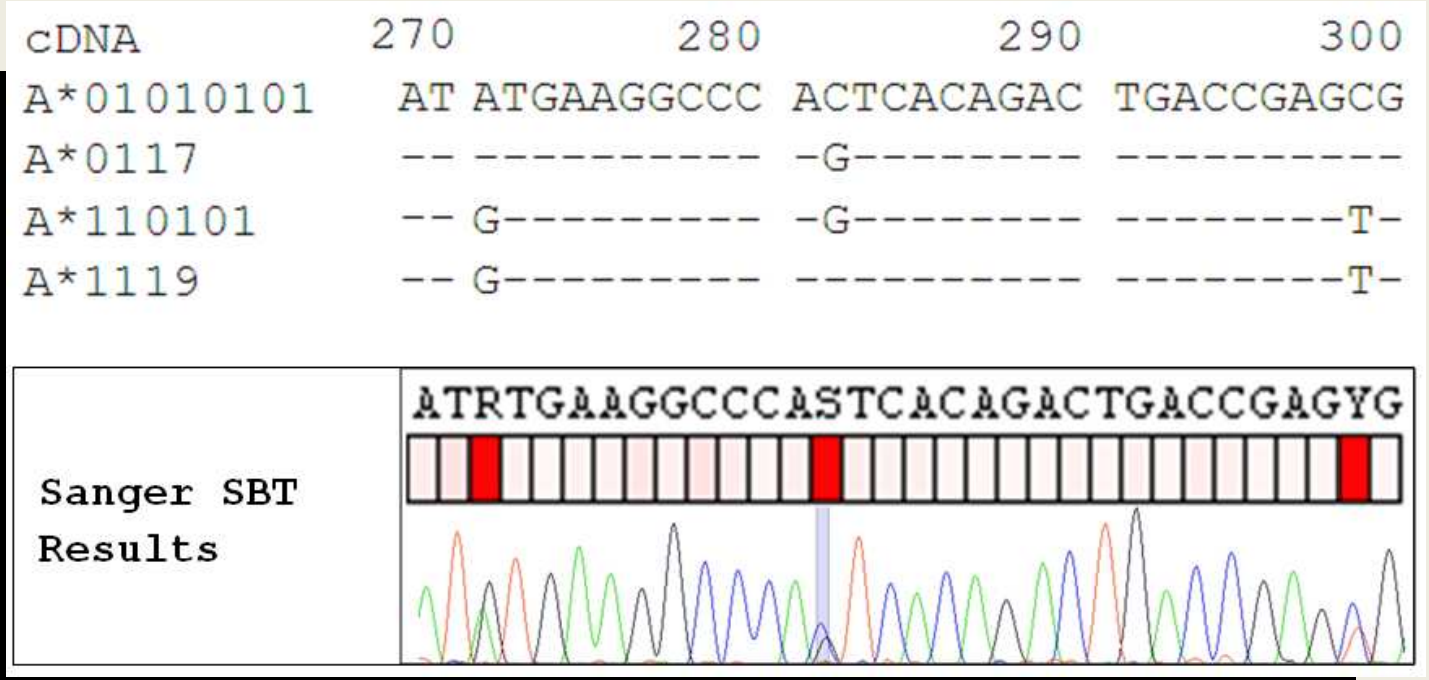
example: HLA-B
B*07:02, 44:02
B*07:02, 44:19N

Ambiguità generate dal sequenziamento dovute a mancanza di determinazione della fase gametica.

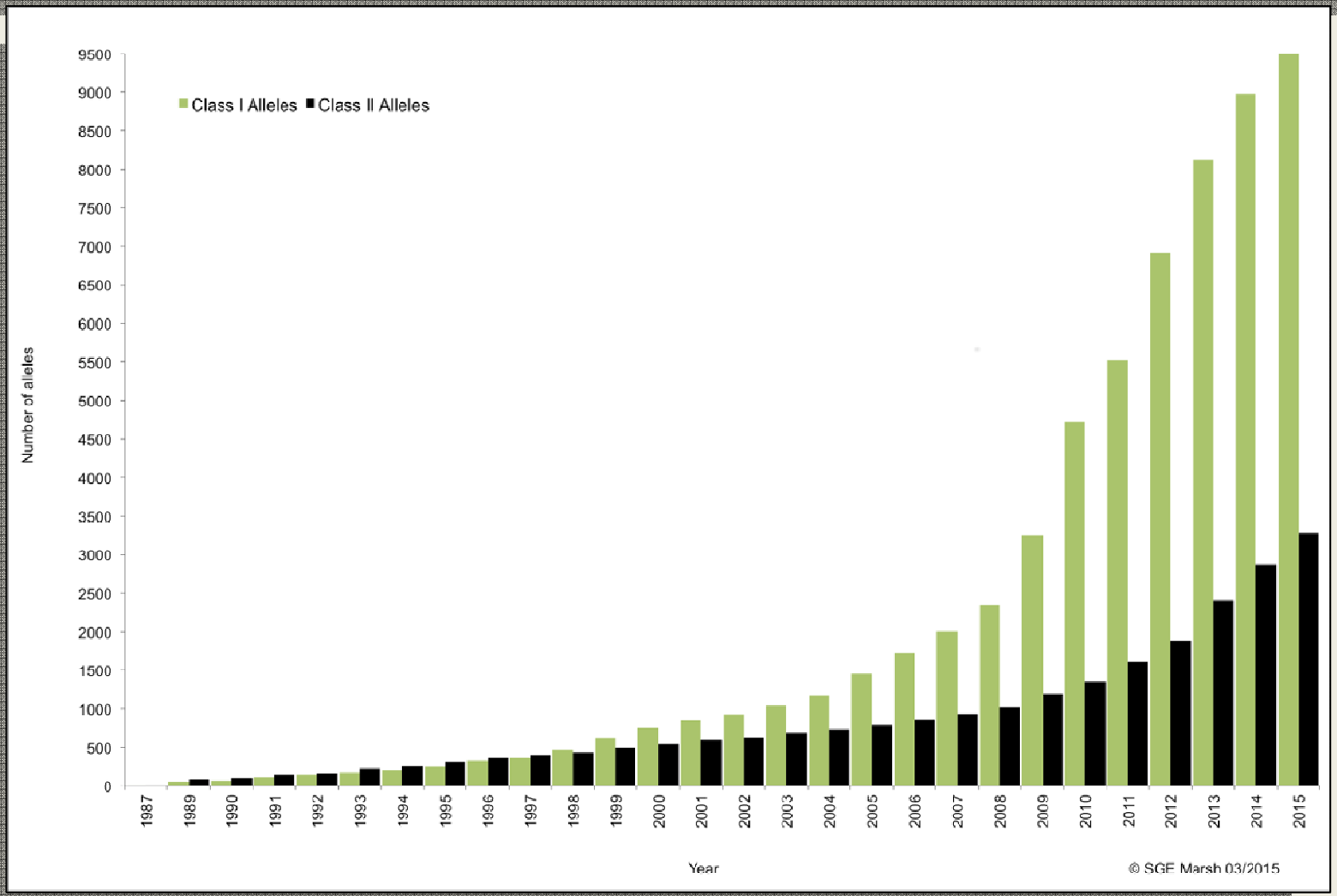
Ambiguità generate dal sequenziamento dovute a mancanza di determinazione della fase gametica.

Allele 1	Allele 2
A*01010101	A*110101
A*0117	A*1119

All
A*
A*0117 A*1119



Increased number of HLA alleles = more ambiguities



The new Gold Standard for genetic test

La Next Generation Sequencing (NGS) è un metodo di sequenziamento del DNA che ha la capacità di processare milioni di frammenti (reads) in parallelo a un costo estremamente ridotto.

Attraverso la NGS un genoma complesso come quello umano può essere sequenziato in meno di 7 giorni a un costo di $ 700-1000.

La NGS è attualmente il metodo di analisi del DNA col più basso tasso di errore per base.

Problemi? A volte il troppo guasta!!

Piattaforme di Next Generation Sequencing (NGS)

- Illumina (Solexa)
  - HiSeq System
  - Genome analyzer IIx
  - MySeq
- Ion Torrent - Life Technologies
  - Personal Genome Machine (PGM)
  - Proton

Next Generation Sequencing
Amplified Single Molecule Sequencing

- Pacific Biosciences
  - PacBio RS
- Oxford Nanopore Technologies
  - GridION System
  - MinION

Third Generation Sequencing,
Next Next Generation Sequencing,
Single Molecule Sequencing

Benchtop genome sequencers

MiSeq

PGM (Personal Genome Machine)

	MiSeq	PGM
lunghezza reads (paia di basi)	300	200
throughput (giga basi G)	15	1
numero di reads per run (milioni)	25	11
Accuratezza (%)	99,90	99,00
Tempo di run (h)	65	4,5

Illumina MiSeq vs Ion Torrent PGM

INTERNATIONAL JOURNAL OF IMMUNOGENETICS

doi: 10.1111/iji.12213

Towards allele-level human leucocyte antigens genotyping – assessing two next-generation sequencing platforms: Ion Torrent Personal Genome Machine and Illumina MiSeq

J. L. Duke*, C. Lind*, K. Mackiewicz*, D. Ferriola*, A. Papazoglou*, O. Derbenevat, D. Wallacet‡ & D. S. Monos*‡

International Journal of Immunogenetics, 2015, 42, 346–358

Apparecchiature associate alla NGS

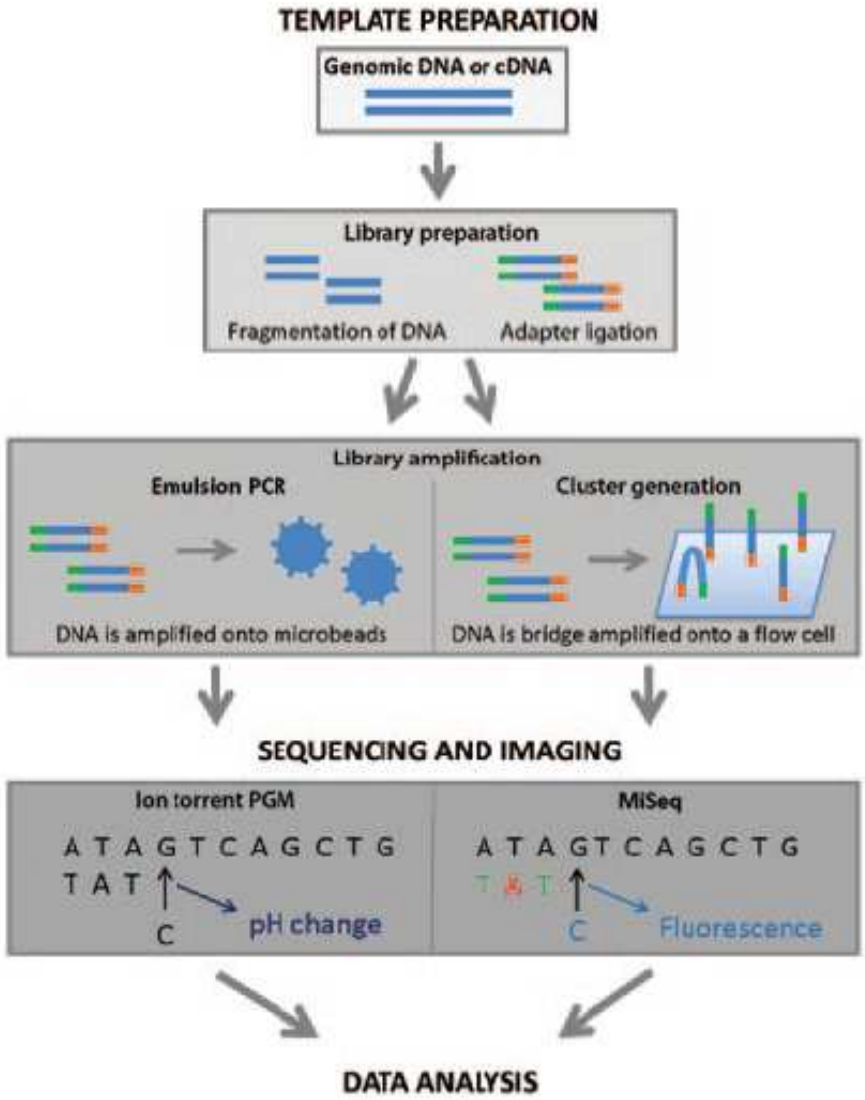
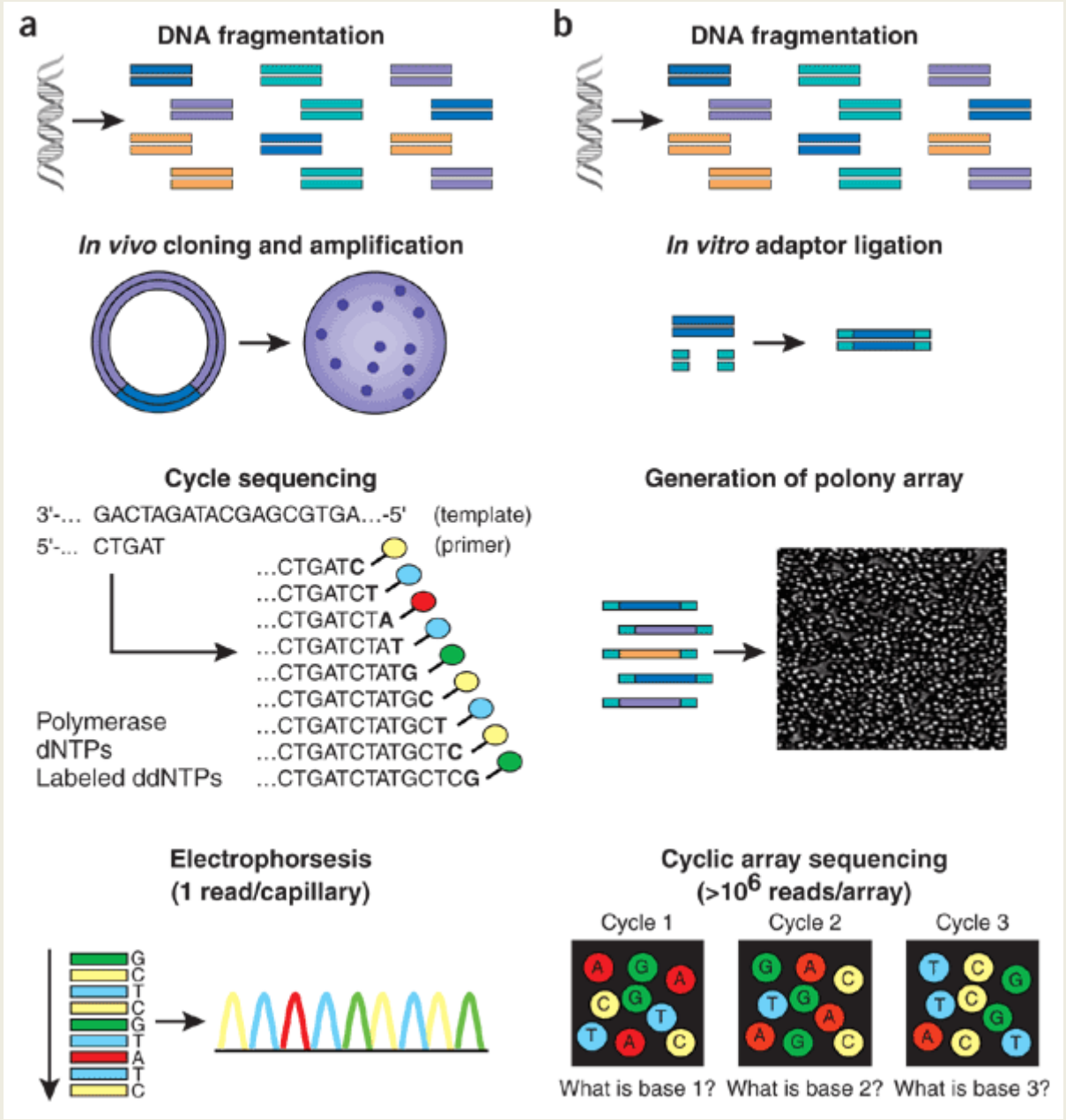
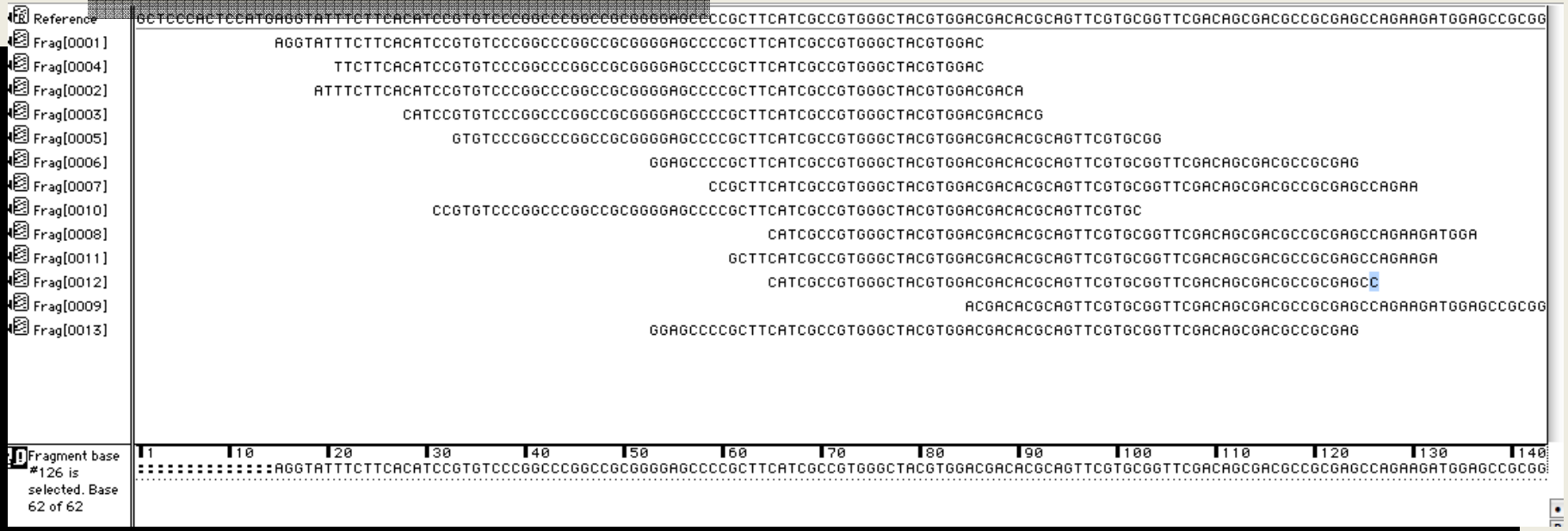


Figure 1. Next-generation sequencing methodology.


a
DNA fragmentation
In vivo cloning and amplification
Cycle sequencing
3'-... GACTAGATACGAGCGTGA...-5' (template)
5'-... CTGAT (primer)
...CTGATC
...CTGATCT
...CTGATCTA
...CTGATCTAT
...CTGATCTATG
...CTGATCTATGC
...CTGATCTATGCT
...CTGATCTATGCTC
...CTGATCTATGCTCG
Polymerase
dNTPs
Labeled ddNTPs
Electrophorsesis
(1 read/capillary)
G
C
T
C
G
T
A
T
C
b
DNA fragmentation
In vitro adaptor ligation
Generation of polony array
Cyclic array sequencing
(>10^6 reads/array)
Cycle 1
Cycle 2
Cycle 3
What is base 1?
What is base 2?
What is base 3?

SOME DEFINITIONS



Read: sequenza ottenuta da un singolo cluster originato a partire da un frammento di DNA o cDNA.

Coverage o Read Depth: numero medio di volte che una base viene analizzata, corrisponde al numero di reads che mappano sulla base.

HLA & NGS

HLA Typing:

- Risoluzione allelica a livello del 6-8 digit
- Alta processività
- Basso costo

La quasi totalità delle tipizzazioni HLA per i registri di donatori è prodotta attualmente con metodi NGS

HLA & NGS

HLA e MALATTIE

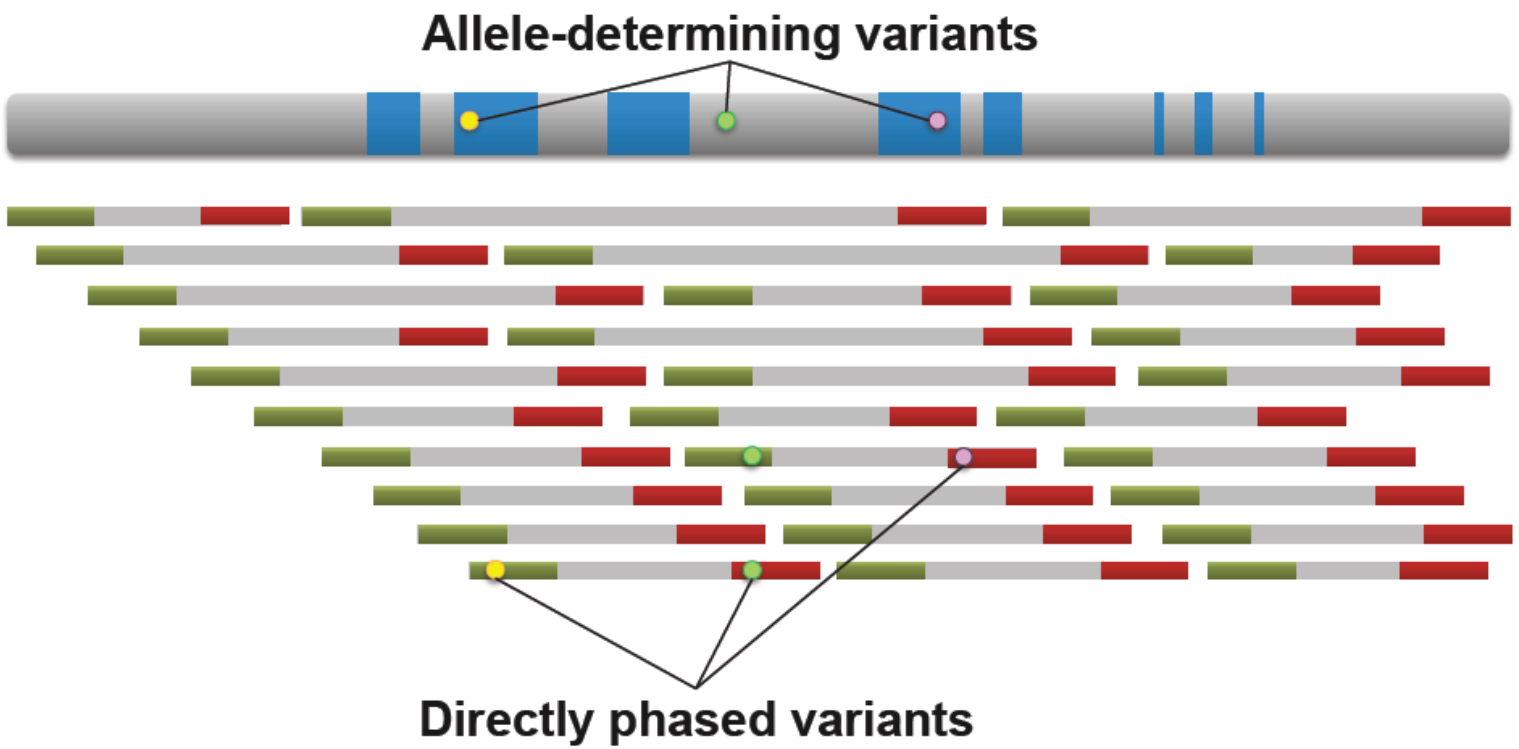
Esistono oltre 100 malattie HLA associate

FARMACOGENOMICA

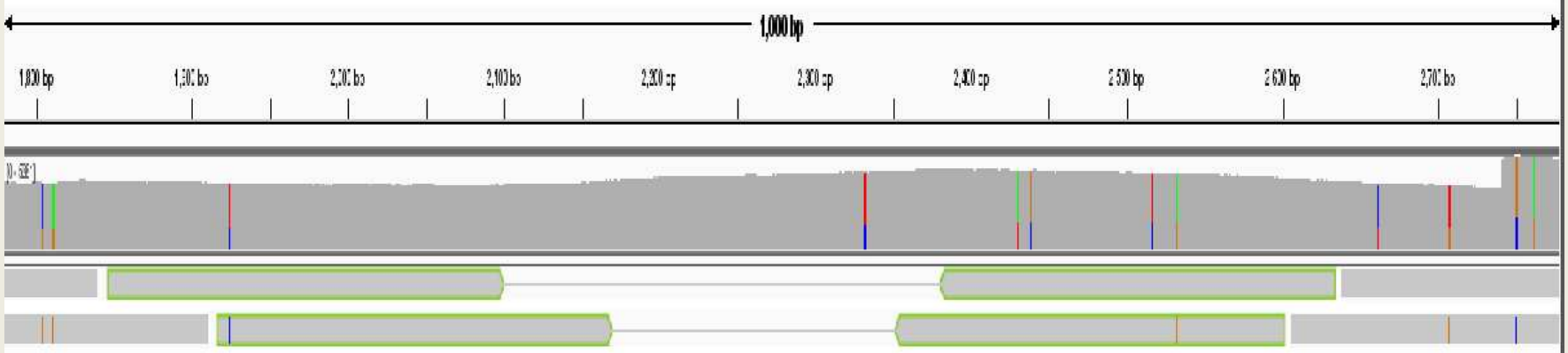
E' crescente l' individuazione di risposte avverse ai farmaci mediate dal specifici alleli HLA.

Con l' NGS sarà possibile indagare sui meccanismi attraverso cui i geni HLA sono modulati nella loro espressione.

TruSight HLA Phase Determination



Phasing Paired-End Sequences



B*07:02:01	T	T	AG	TA
B*41:01	C	C	TC	CG

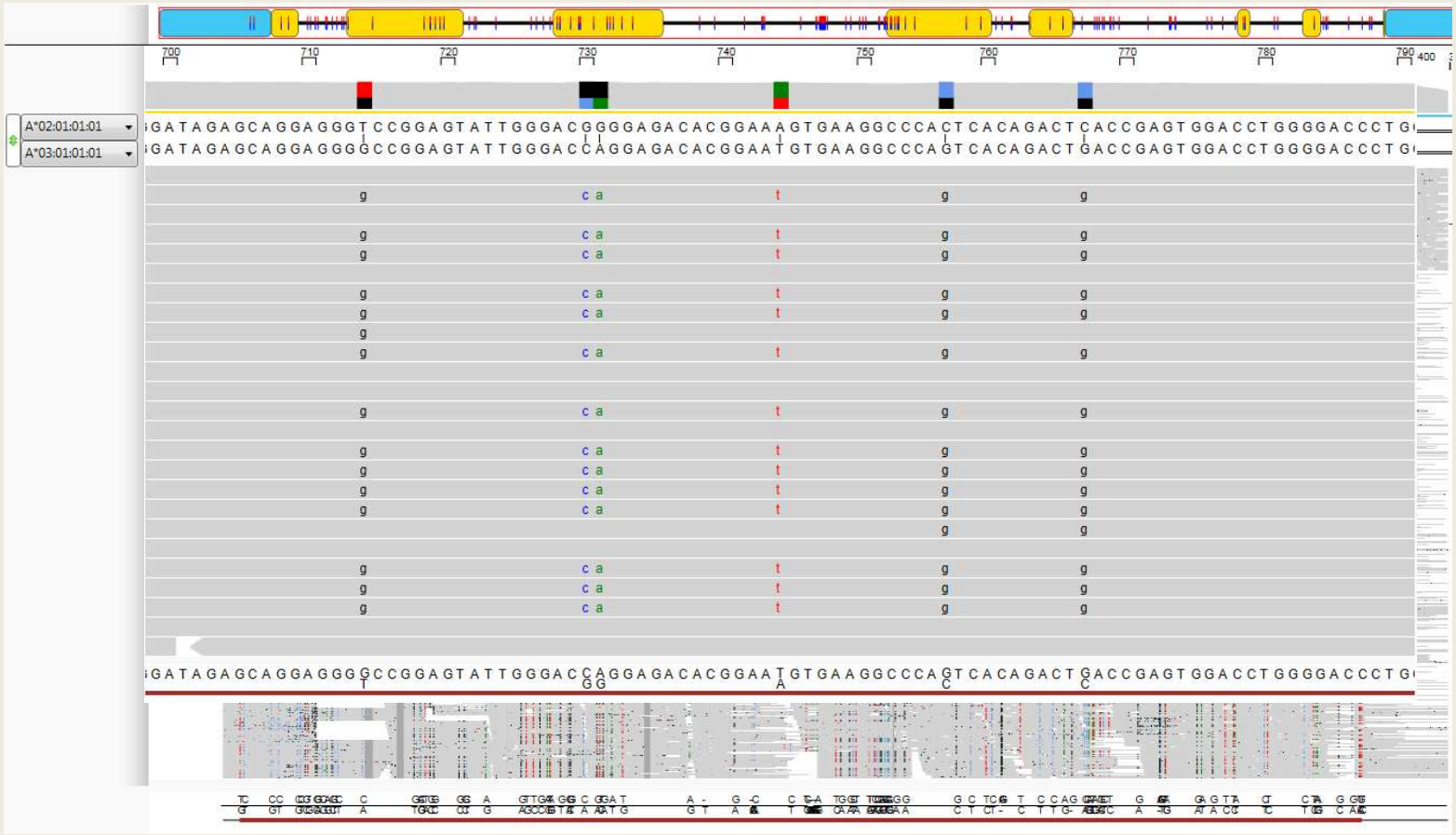
Insert sizes: 787 and 685 bases
Phased Polymorphisms 609 bases apart
Using paired-end 250 base reads

Allele 1 **Allele 2**

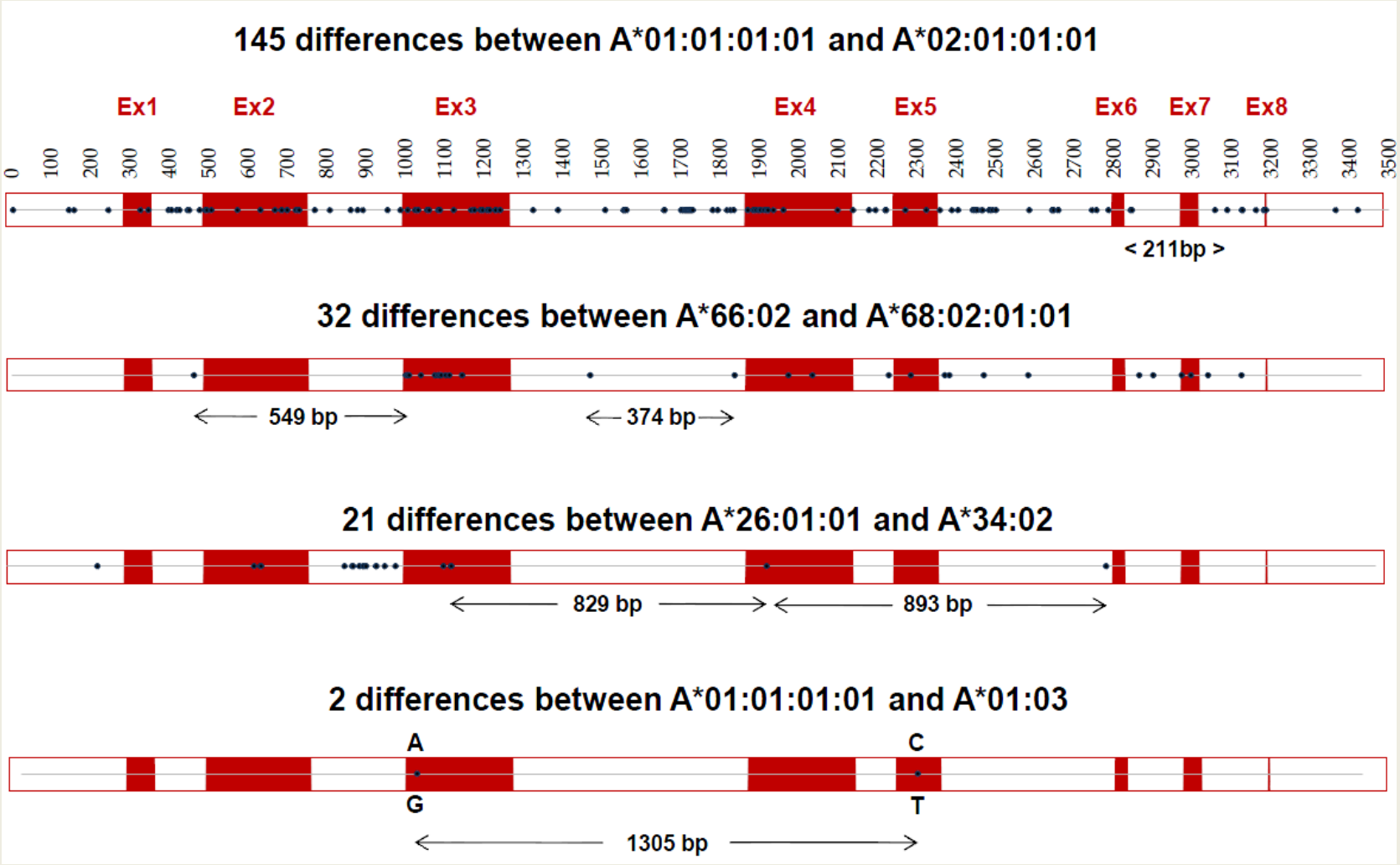
A*02:01:01:01 (CWD) A*03:01:01:01 (CWD)

A*02:35:01 (CWD) A*03:108 (WD)

```
cDNA                230       240       250       260       270       280       290       300
A*02:01:01:01       GATAGA GCAGGAGGGT CCGGAGTATT GGGACGGGGA GACACGGAAA GTGAAGGCCC ACTCACAGAC TCACCGAGTG
A*02:35:01          ------ ---------- ---------- ---------- ---------- ---------- -G-------- -G--------
A*03:01:01:01       ------ ---------G ---------- -----CA--- ---------T ---------- -G-------- -G--------
A*03:108            ------ ---------G ---------- -----CA--- ---------T ---------- ---------- ----------
```

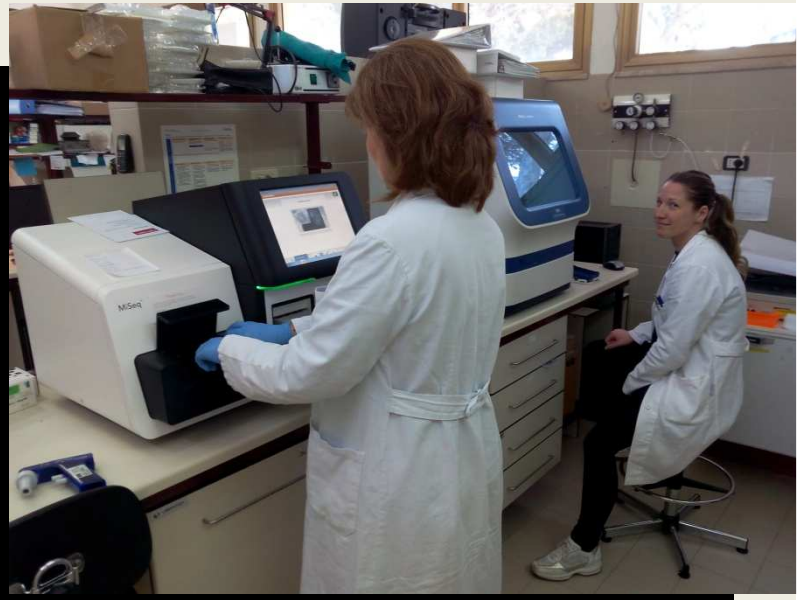


The Significance of Read Length for HLA Typing by NGS



Lind C et al., Hum Immunol. 10:1033-42 (2010)

Applicazione di un metodo di high-troughput basato sulla Next Generation Sequencing per la tipizzazione dei loci HLA



Per la validazione del metodo sono stati identificati 96 campioni di DNA , precedentemente tipizzati mediante SBT ad alta risoluzione.

Kit commerciali (32 Campioni con GENDX NGSgo; 20 con OMIXON Holotype HLA)

288 con un sitema Home-made individual tagging sequencing method

Nel protocollo di validazione abbiamo incluso i kit commerciali

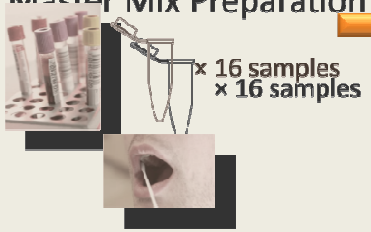
GENDX

omixon

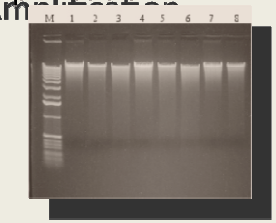
Steps in Holotype HLA

5 hours hands-on time

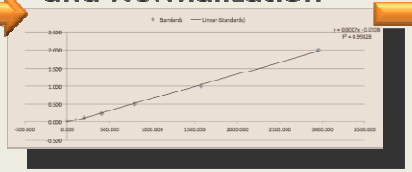
Step 1
HLA Amplification
Master Mix Preparation
0h 15 hands on time
0h 15 total time



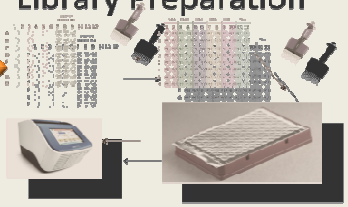
Step 2
HLA Class I and Class II
Amplification
1h 30 hands on time
8h 15 total time



Step 3
Amplicon Quantitation
and Normalization
1h 0 hands on time
1h 0 total time

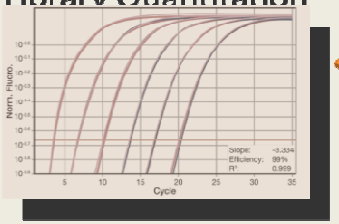


Step 4
Library Preparation
1h 10 hands on time
3h 15 total time



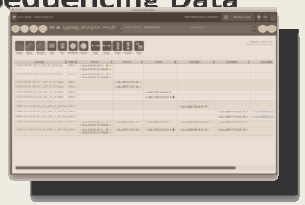
Step 5
Library Size Selection
0h 15 hands on time
1h 0 total time

Step 6
Library Quantitation
0h 5 hands on time
1h 0 total time



Step 7
Sequencing on Illumina
MiSeq
1h 0 hands on time
40h 30 total time

Step 8
Anaysis of HLA
Sequencing Data
0h 0 hands on time
6h 0 total time



Overview Illumina Workflow

GenDx offers with NGSgo® a full workflow for the Illumina MiSeq platform, from amplification, library preparation to data analysis. We made it easy for you to make a headstart by integrating reagents, software and our know-how in this fully integrated workflow. Together we enter this next level of NGS-HLA typing.

Our Method

Long Range PCR specifica per i loci HLA-A, B, C, DRB1, DQB1, DPB1, basata su piattaforma Illumina Nextera XT e sequenziamento su Illumina MiSeq

Sviluppo di un protocollo di liquid handling semi-automatico, su MTP 96

Sviluppo di un supporto informatico Home Made.

Perché un metodo Home-made?

Perché un metodo Home-made?

La NGS è un metodo complesso con contenuti importanti di genomica e bioinformatica.

La NGS è un metodo complesso con contenuti importanti di genomica e bioinformatica.

Interpretazione dei dati finali più confidente, soprattutto nel casi in cui il software produce un risultato dubbio.

Interpretazione dei dati finali più confidente, soprattutto nel casi in cui il software produce un risultato dubbio.

Permette di ottenere un *know how* utile a sviluppare altri sistemi di analisi genetica (loci KIR)

Permette di ottenere un know how utile a sviluppare altri sistemi di analisi genetica (loci KIR)

Permette di trovare soluzioni specifiche a problemi specifici.

Permette di trovare soluzioni specifiche a problemi specifici.

Hosomichi *et al. BMC Genomics* 2013, **14**:355
http://www.biomedcentral.com/1471-2164/14/355

BMC Genomics

METHODOLOGY ARTICLE **Open Access**

Phase-defined complete sequencing of the HLA genes by next-generation sequencing

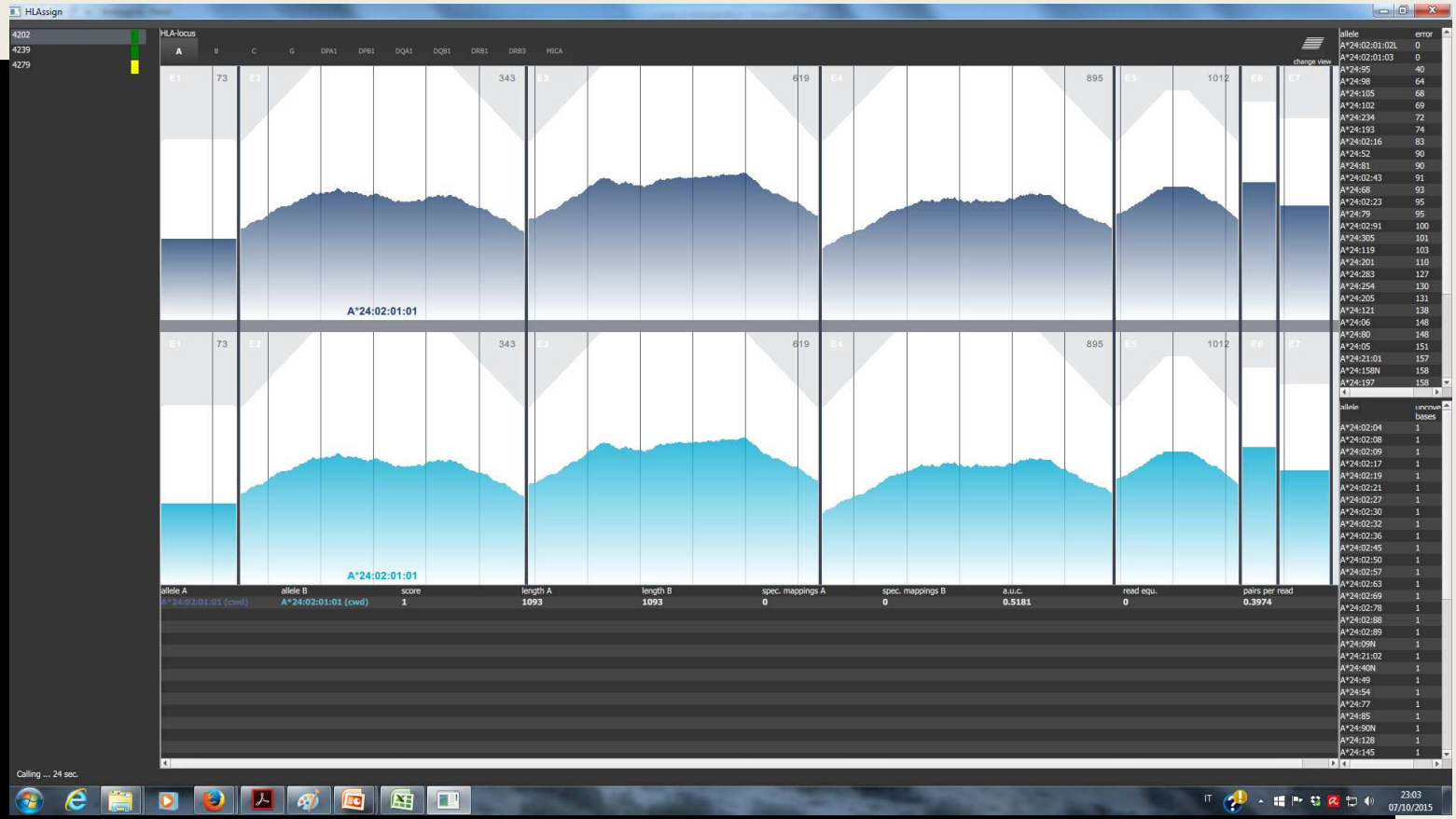
Kazuyoshi Hosomichi[1], Timothy A Jinam[1], Shigeki Mitsunaga[2], Hirofumi Nakaoka[1] and Ituro Inoue[1*]

- 7 Prodotti di LongRange PCR dal 5' al 3' UTR per i loci: *A, B, C*, DRB1, DPB1, DQB1, *G* per un target di ~ 50Kb
- Preparazione delle librerie con kit Nextera XT modificato e individual tag
- Selezione dei frammenti mediante doppio taglio con AmpureXP Beads
- Normalizzazione delle librerie con Nextera XT normalization module
- Sequenziamento su Illumina MiSeq con Flow Cell v3 300+300 cicli

Nucleic Acids Research, 2015 **1**
doi: 10.1093/nar/gkv184

Development of a high-resolution NGS-based HLA-typing and analysis pipeline

Michael Wittig[1,*], Jarl A. Anmarkrud[2,3,4], Jan C. Kässens[5], Simon Koch[6], Michael Forster[1], Eva Ellinghaus[1], Johannes R. Hov[2,3,4,7], Sascha Sauer[8], Manfred Schimmler[5], Malte Ziemann[9], Siegfried Görg[9], Frank Jacob[6], Tom H. Karlsen[2,3,4,7,*] and Andre Franke[1,*]

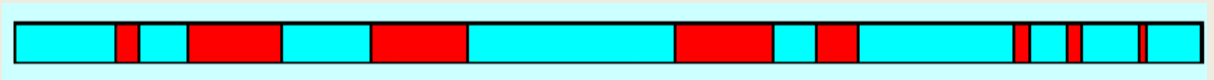


Strategies for HLA Typing by NGS

Exon-based typing (selected exonic/intronic sequences)

cDNA-based typing (selected exons can be included)

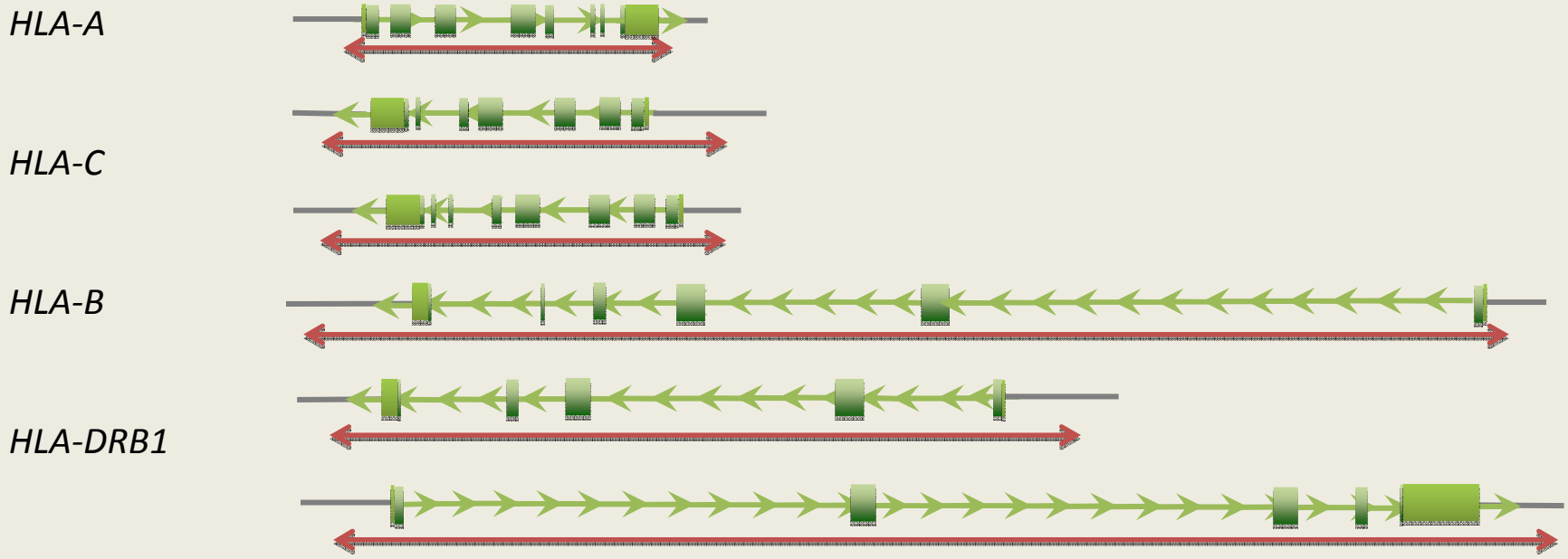
Full genomic typing



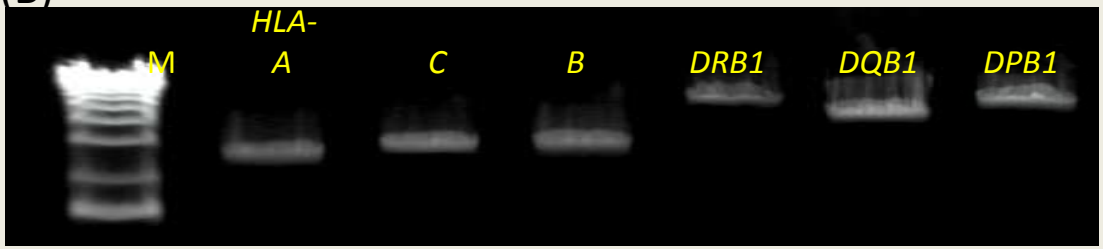
Full length Amplification (5'UTR to 3'UTR) of : HLA-A, HLA-B, HLA-C (~3kb)
HLA-DQA1 (~7kb)
HLA-DQB1 (~7kb)

Partial Amplification of: **HLA-DRB1 amplicon ~5 kb (gene ~15kb)**
HLA-DPB1 amplicon ~7kb (gene ~12kb)

(A)



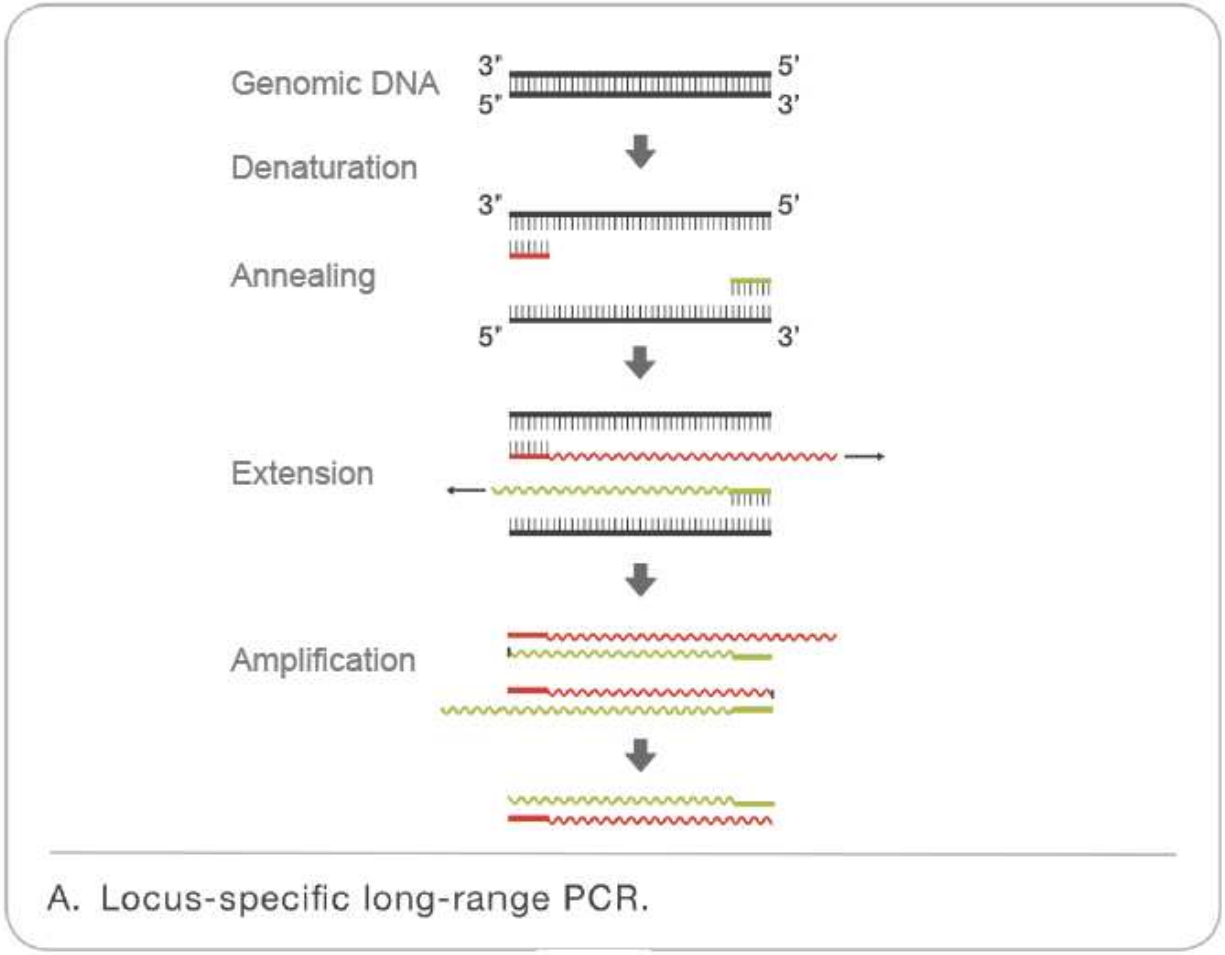
(B)



Locus	length (bp)
HLA-A	3,398
HLA-C	4,296
HLA-B	4,440
HLA-DRB1	11,899
HLA-DQB1	7,118
HLA-DPB1	13,605

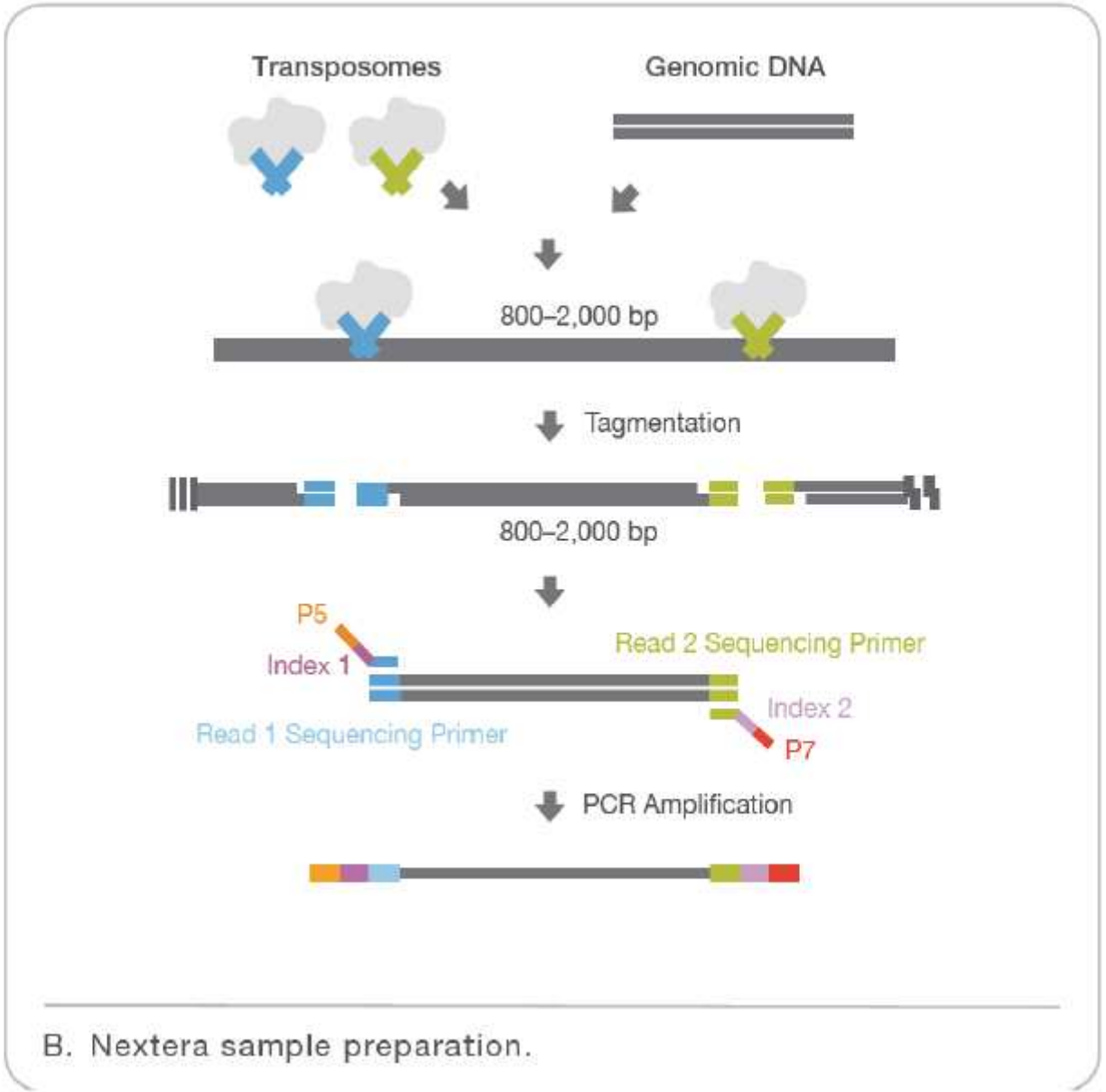
Figure S1: PCR amplification of the six HLA genes. (A) Amplified region of the six loci, where dark green boxes represent exons. Red arrows indicate the amplified region. (B) Agarose gel electrophoresis of PCR products and sizes of each amplicon.

Step 1: Amplification

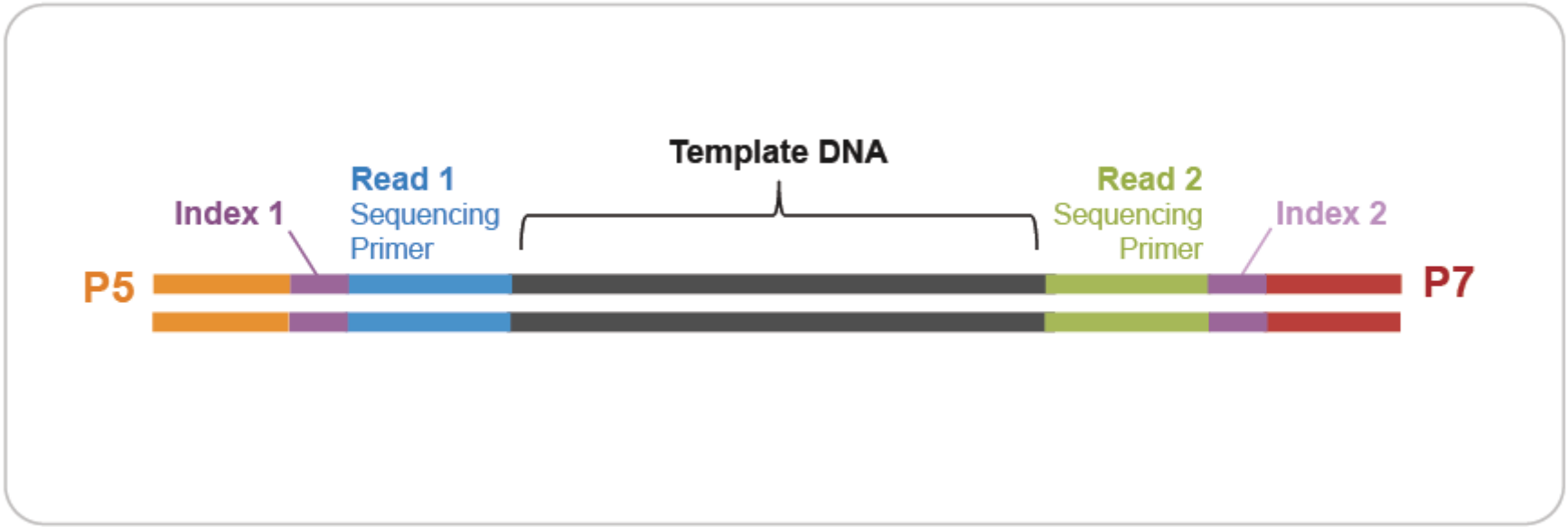


A. Locus-specific long-range PCR.

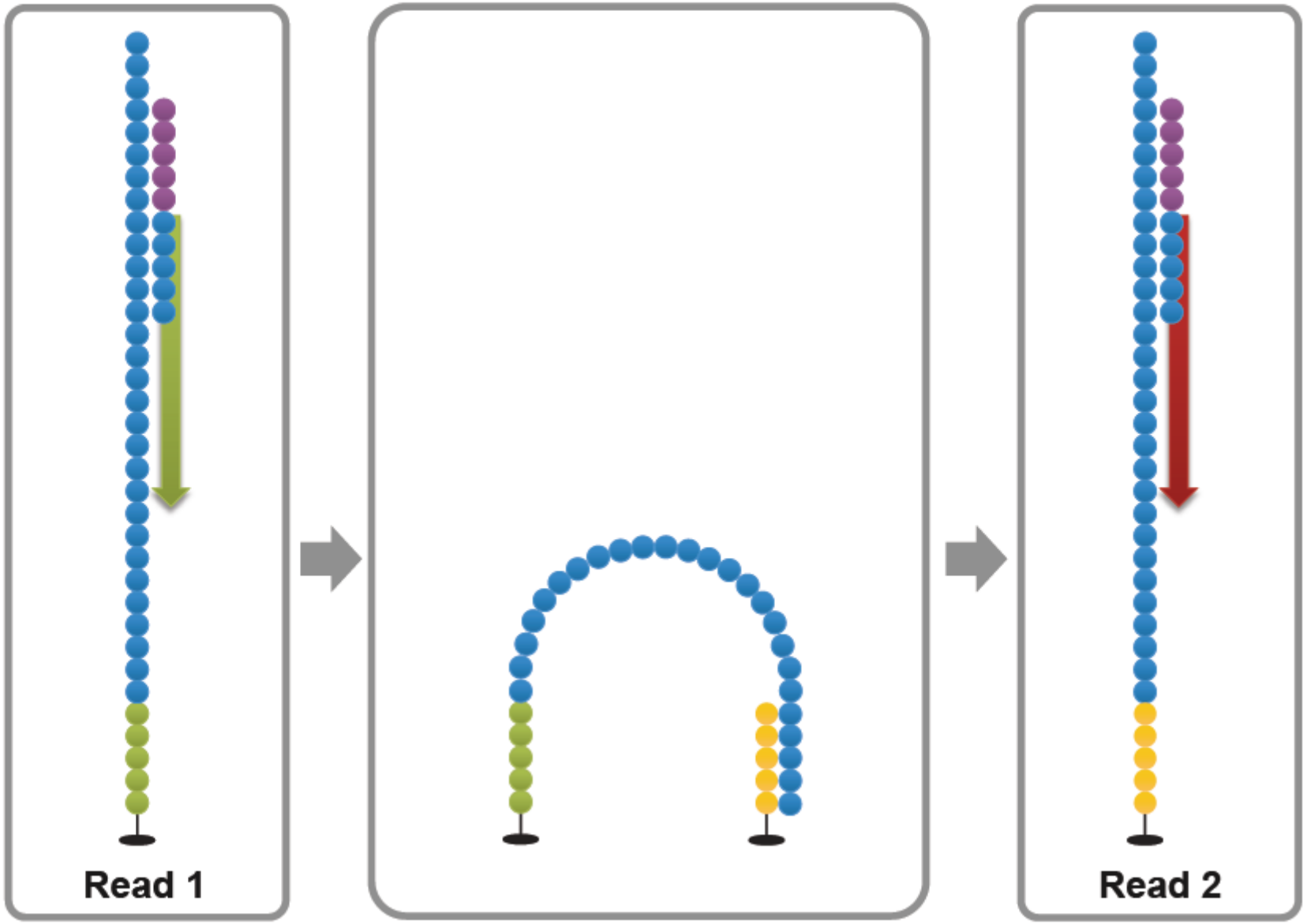
Step 2: Nextera Library Prep

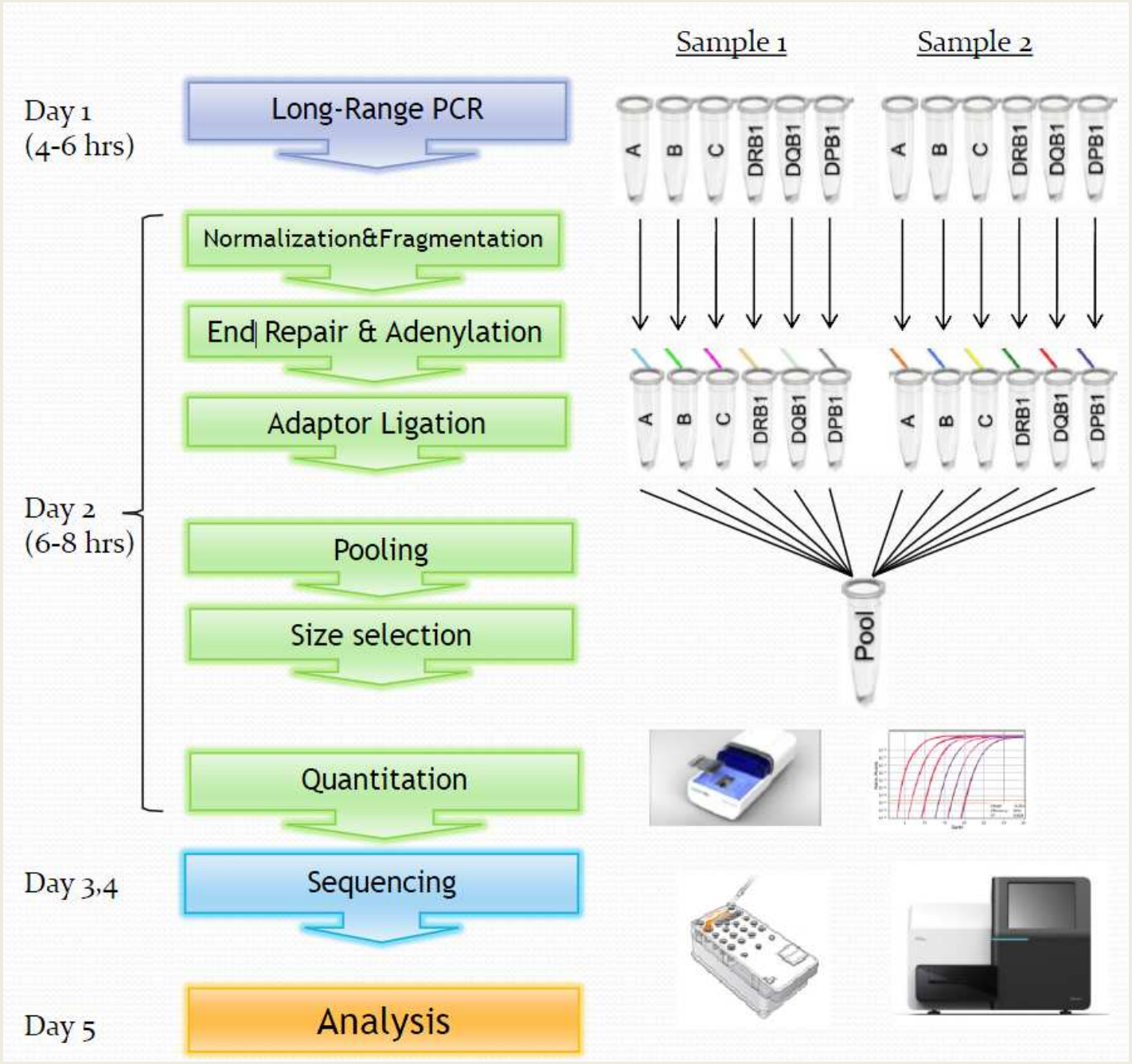


B. Nextera sample preparation.


Template DNA
Read 1
Sequencing
Primer
Read 2
Sequencing
Primer
Index 1
Index 2
P5
P7

Illumina Paired-End Sequencing




Sample 1
Sample 2
Day 1
(4-6 hrs)
Long-Range PCR
A
B
C
DRB1
DQB1
DPB1
Normalization&Fragmentation
End Repair & Adenylation
Adaptor Ligation
Day 2
(6-8 hrs)
Pooling
Size selection
Pool
Quantitation
Day 3,4
Sequencing
Day 5
Analysis

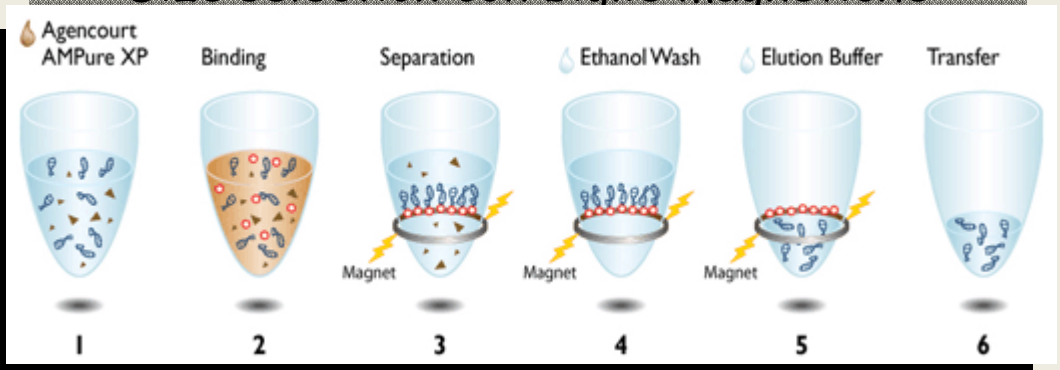
Workflow

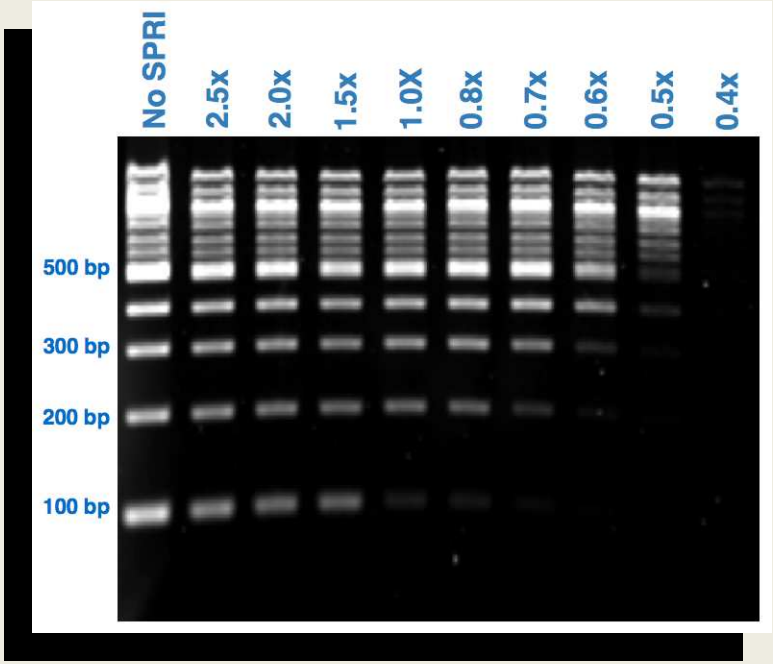
(tempi riferiti a una plate da 96 campioni)

1. Amplificazione O/N dei loci d'interesse — 1° giorno
2. Dosaggio dei frammenti e produzione di un pool di prodotti equimolari
3. Dosaggio fluorimetrico e diluizione intermedia dei pool — 2° giorno
4. Dosaggio fluorimetrico e diluizione di lavoro dei pool
5. Tagmentazione e purificazione
6. Indexing PCR
7. Doppia purificazione per la selezione dei frammenti — 3° giorno
8. Normalizzazione delle librerie
9. Produzione del pool di librerie equimolari
10. Sequenziamento su Illumina MiSeq con Flow Cell v3 301+301 cicli — 56h
11. Analisi dei dati

La preparazione delle librerie

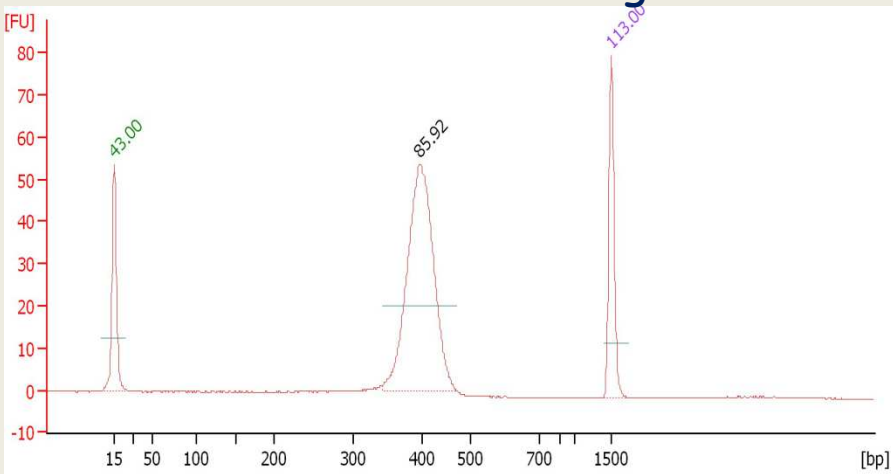
Size selection con biglie magnetiche





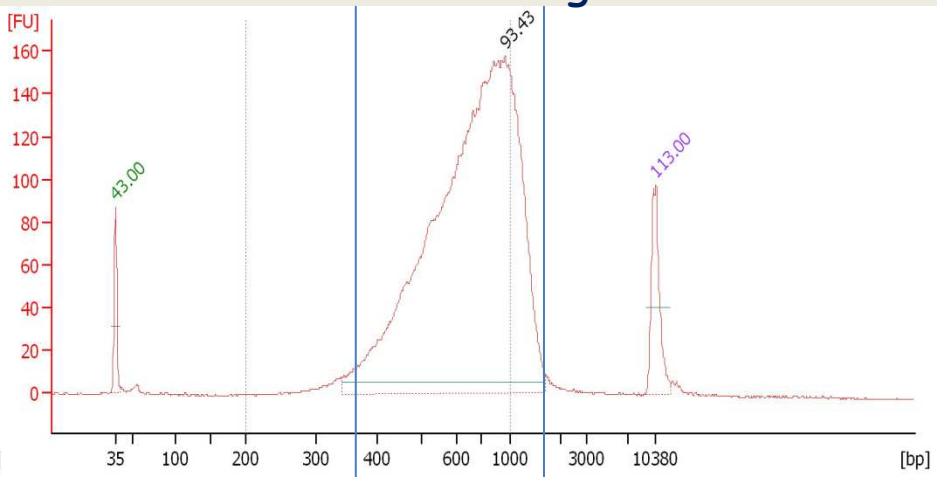
La preparazione delle librerie

Selezione da gel



Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		396	8.88	34.0	
3	▶	1,500	2.10	2.1	Upper Marker

Selezione con biglie



Peak		Size [bp]	Conc. [pg/µl]	Molarity [pmol/l]	Observations
1	◀	35	125.00	5,411.3	Lower Marker
2		940	1,790.38	2,886.1	
3	▶	10,380	75.00	10.9	Upper Marker

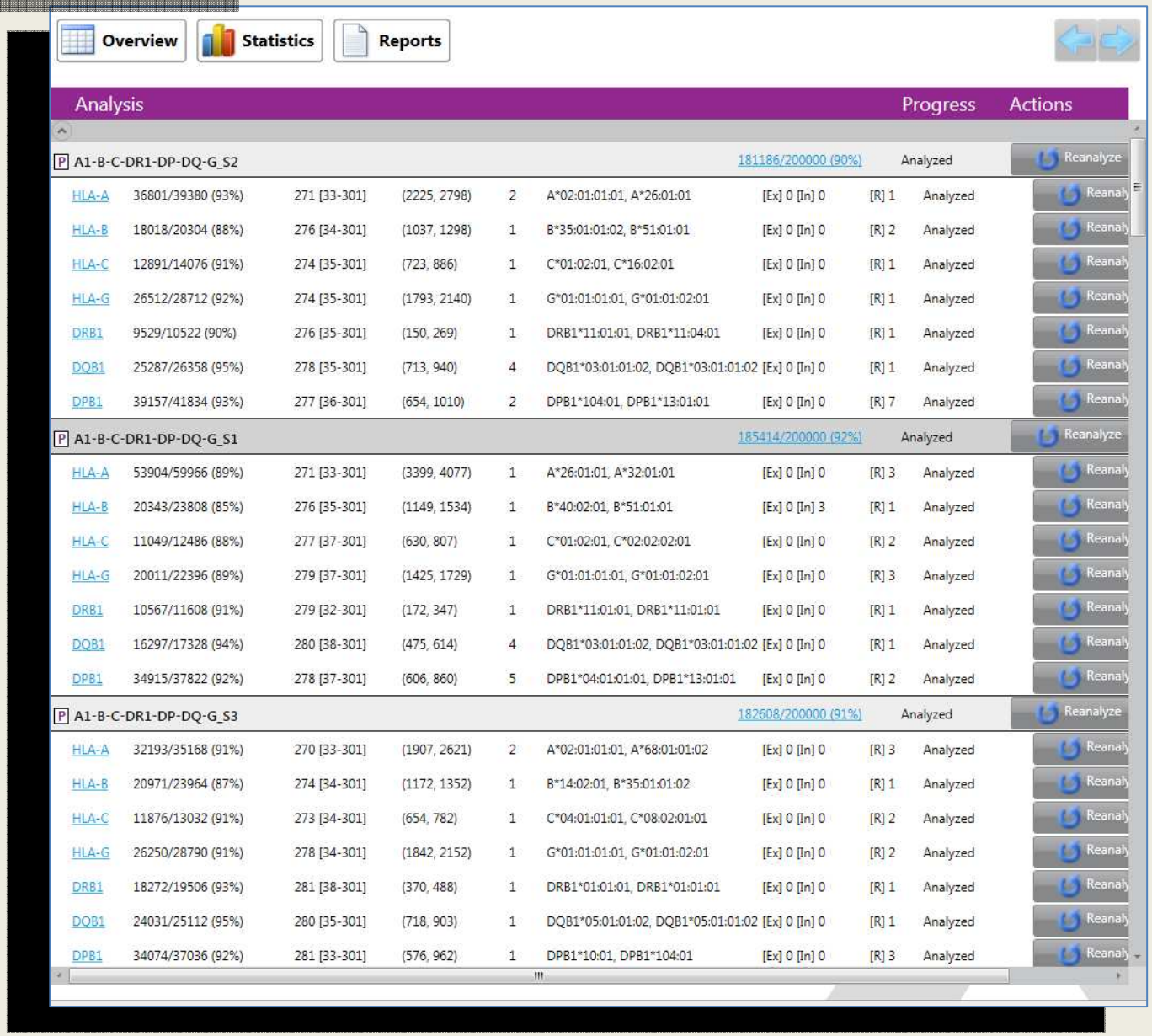
2° step, beads 0,6X
lega frammenti più piccoli

Discard supernatant
Keep beads

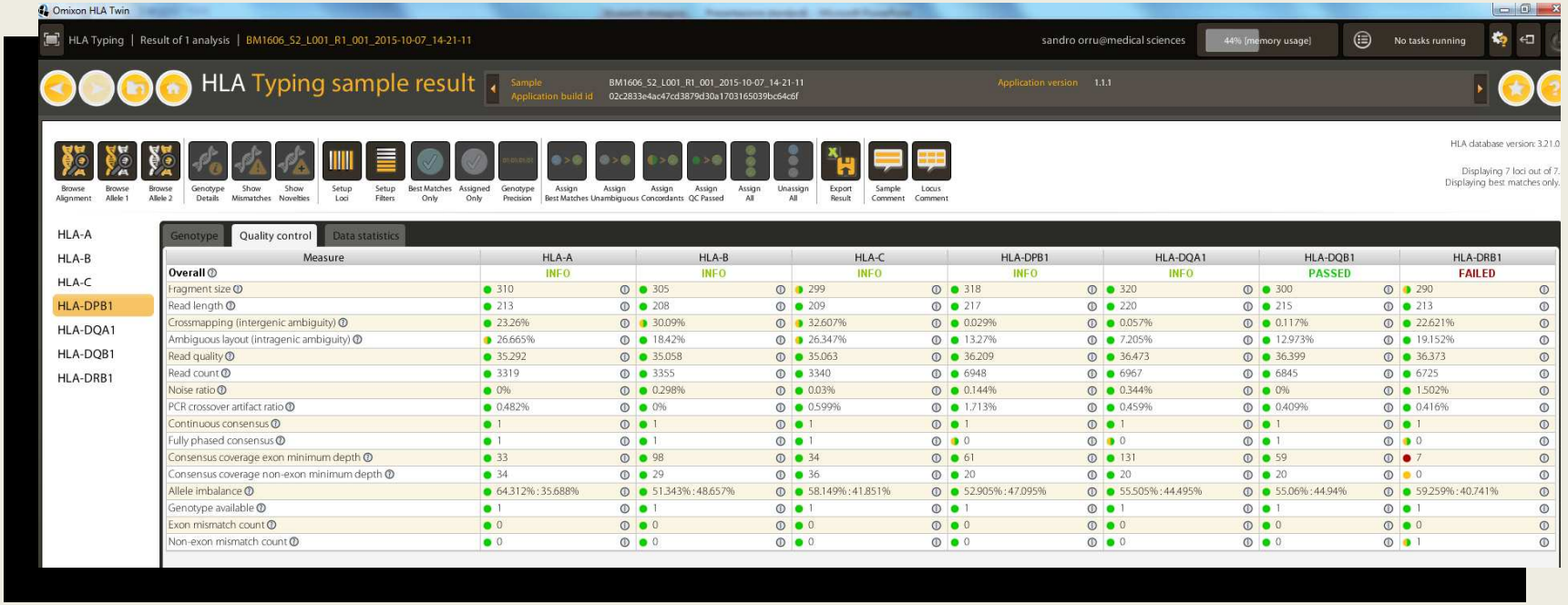
1° step, beads 0,5X
lega frammenti grandi

Keep supernatant
Discard beads

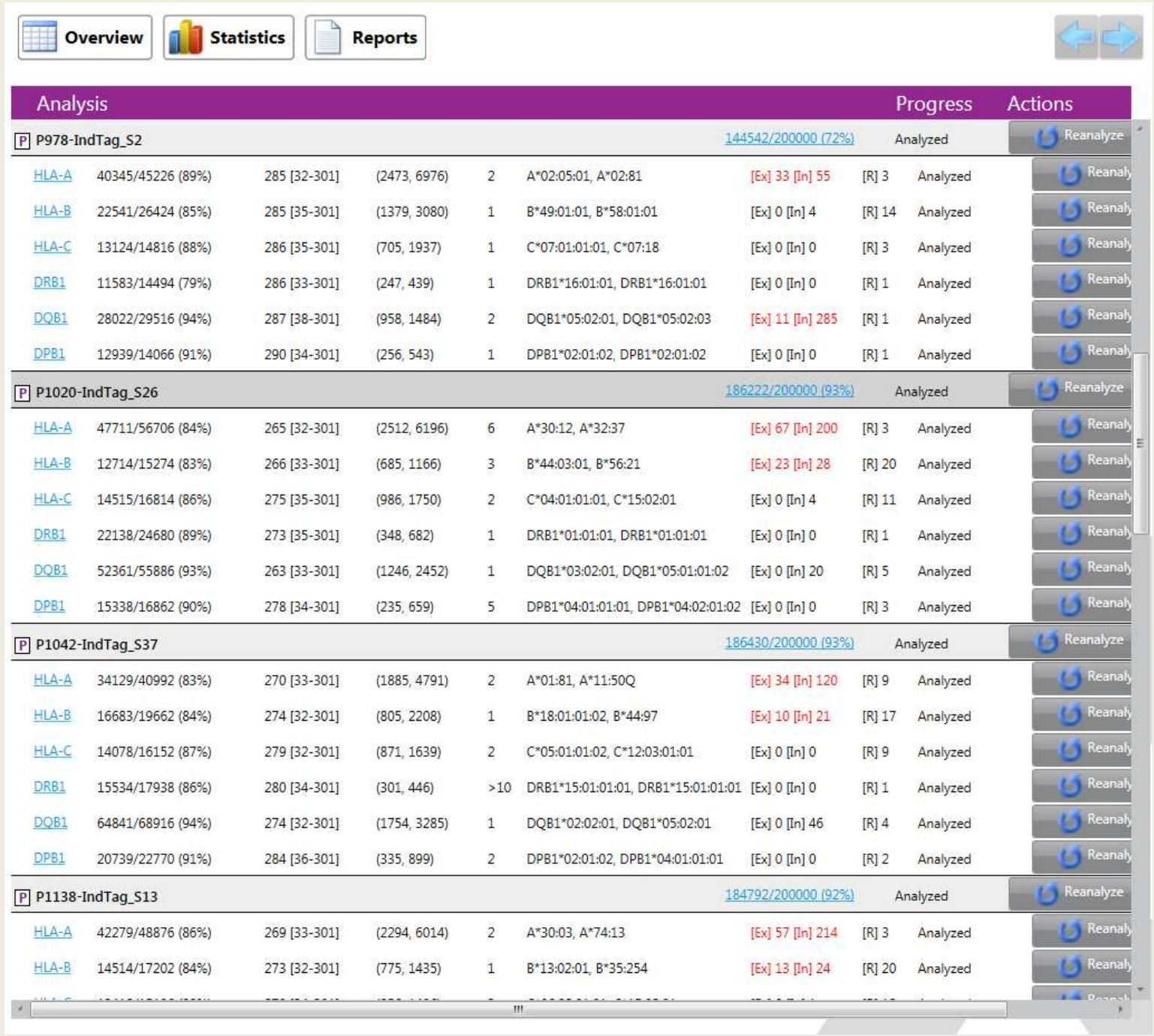
NGSengine GenDX



HLA Twin Omixon

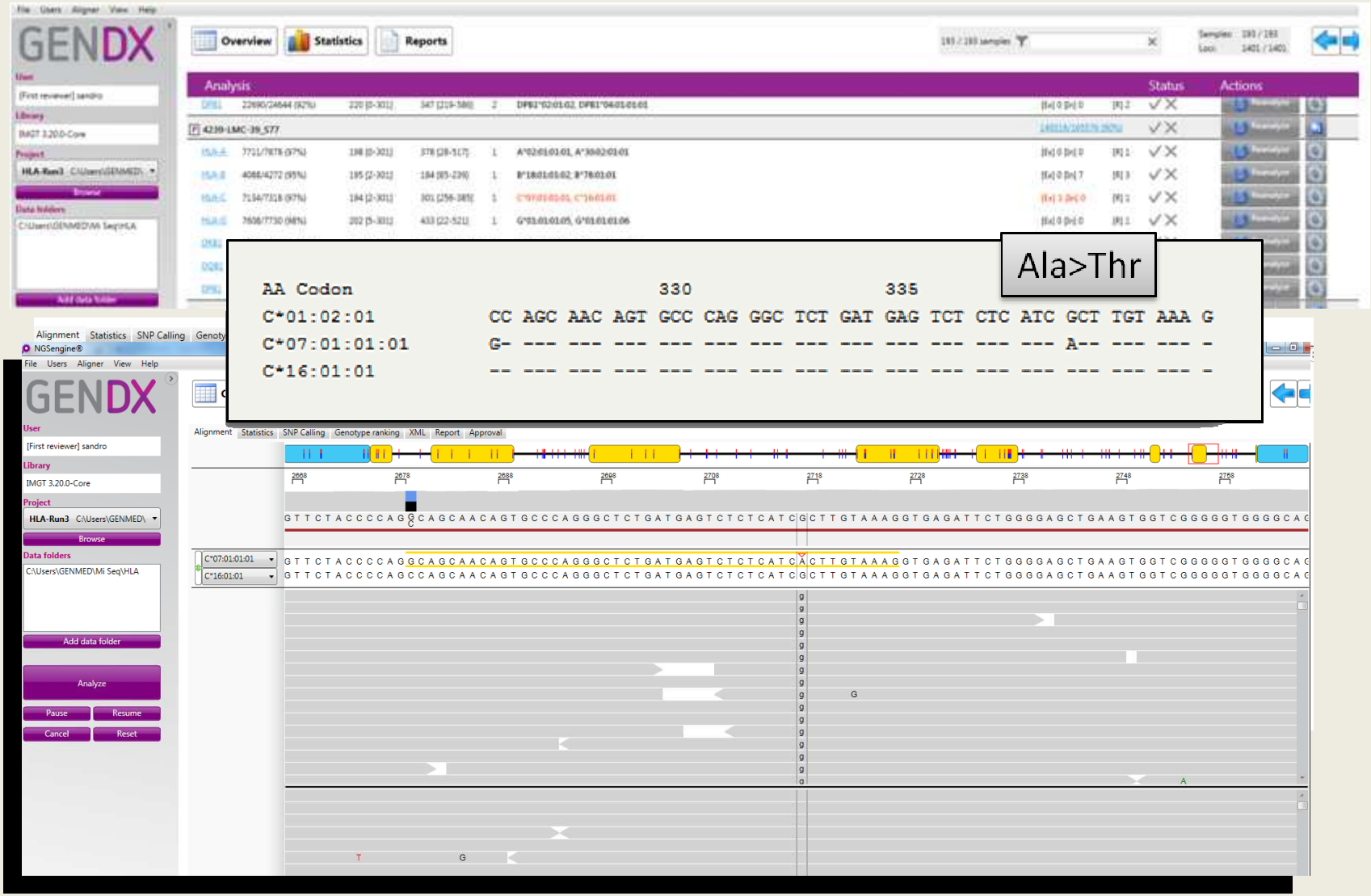



Overview
A1-B-C-DR1-DP-DQ-GS14 HLA-A
Alignment
Statistics
SNP Calling
Allele ranking
Genotype ranking
XML
Report
Approval
A*02:01:01:01
A*26:01:01
gDNA Position:
931
A: 5 , C: 995 , G: 887 , T: 6 , -: 0
Rejections
Separate alleles
Show library
Show reads


Overview
Statistics
Reports
Analysis
Progress
Actions
P978-IndTag_S2
144542/200000 (72%)
Analyzed
Reanalyze
HLA-A 40345/45226 (89%) 285 [32-301] (2473, 6976) 2 A*02:05:01, A*02:81 [Ex] 33 [In] 55 [R] 3 Analyzed
HLA-B 22541/26424 (85%) 285 [35-301] (1379, 3080) 1 B*49:01:01, B*58:01:01 [Ex] 0 [In] 4 [R] 14 Analyzed
HLA-C 13124/14816 (88%) 286 [35-301] (705, 1937) 1 C*07:01:01:01, C*07:18 [Ex] 0 [In] 0 [R] 3 Analyzed
DRB1 11583/14494 (79%) 286 [33-301] (247, 439) 1 DRB1*16:01:01, DRB1*16:01:01 [Ex] 0 [In] 0 [R] 1 Analyzed
DQB1 28022/29516 (94%) 287 [38-301] (958, 1484) 2 DQB1*05:02:01, DQB1*05:02:03 [Ex] 11 [In] 285 [R] 1 Analyzed
DPB1 12939/14066 (91%) 290 [34-301] (256, 543) 1 DPB1*02:01:02, DPB1*02:01:02 [Ex] 0 [In] 0 [R] 1 Analyzed
P1020-IndTag_S26
186222/200000 (93%)
Analyzed
Reanalyze
HLA-A 47711/56706 (84%) 265 [32-301] (2512, 6196) 6 A*30:12, A*32:37 [Ex] 67 [In] 200 [R] 3 Analyzed
HLA-B 12714/15274 (83%) 266 [33-301] (685, 1166) 3 B*44:03:01, B*56:21 [Ex] 23 [In] 28 [R] 20 Analyzed
HLA-C 14515/16814 (86%) 275 [35-301] (986, 1750) 2 C*04:01:01:01, C*15:02:01 [Ex] 0 [In] 4 [R] 11 Analyzed
DRB1 22138/24680 (89%) 273 [35-301] (348, 682) 1 DRB1*01:01:01, DRB1*01:01:01 [Ex] 0 [In] 0 [R] 1 Analyzed
DQB1 52361/55886 (93%) 263 [33-301] (1246, 2452) 1 DQB1*03:02:01, DQB1*05:01:01:02 [Ex] 0 [In] 20 [R] 5 Analyzed
DPB1 15338/16862 (90%) 278 [34-301] (235, 659) 5 DPB1*04:01:01:01, DPB1*04:02:01:02 [Ex] 0 [In] 0 [R] 3 Analyzed
P1042-IndTag_S37
186430/200000 (93%)
Analyzed
Reanalyze
HLA-A 34129/40992 (83%) 270 [33-301] (1885, 4791) 2 A*01:81, A*11:50Q [Ex] 34 [In] 120 [R] 9 Analyzed
HLA-B 16683/19662 (84%) 274 [32-301] (805, 2208) 1 B*18:01:01:02, B*44:97 [Ex] 10 [In] 21 [R] 17 Analyzed
HLA-C 14078/16152 (87%) 279 [32-301] (871, 1639) 2 C*05:01:01:02, C*12:03:01:01 [Ex] 0 [In] 0 [R] 9 Analyzed
DRB1 15534/17938 (86%) 280 [34-301] (301, 446) >10 DRB1*15:01:01:01, DRB1*15:01:01:01 [Ex] 0 [In] 0 [R] 1 Analyzed
DQB1 64841/68916 (94%) 274 [32-301] (1754, 3285) 1 DQB1*02:02:01, DQB1*05:02:01 [Ex] 0 [In] 46 [R] 4 Analyzed
DPB1 20739/22770 (91%) 284 [36-301] (335, 899) 2 DPB1*02:01:02, DPB1*04:01:01:01 [Ex] 0 [In] 0 [R] 2 Analyzed
P1138-IndTag_S13
184792/200000 (92%)
Analyzed
Reanalyze
HLA-A 42279/48876 (86%) 269 [33-301] (2294, 6014) 2 A*30:03, A*74:13 [Ex] 57 [In] 214 [R] 3 Analyzed
HLA-B 14514/17202 (84%) 273 [32-301] (775, 1435) 1 B*13:02:01, B*35:254 [Ex] 13 [In] 24 [R] 20 Analyzed


Overview
P978-IndTagS2
HLA-A
Alignment
Statistics
SNP Calling
Allele ranking
Genotype ranking
XML
Report
Approval
A*02:05:01
A*02:81
gDNA Position:
656
A: 143 , C: 1 , G: 280 , T: 1 , -: 2
Rejections
Separate alleles
Show library
Show reads

Identificazione di nuovi alleli by GenDX NGSengine



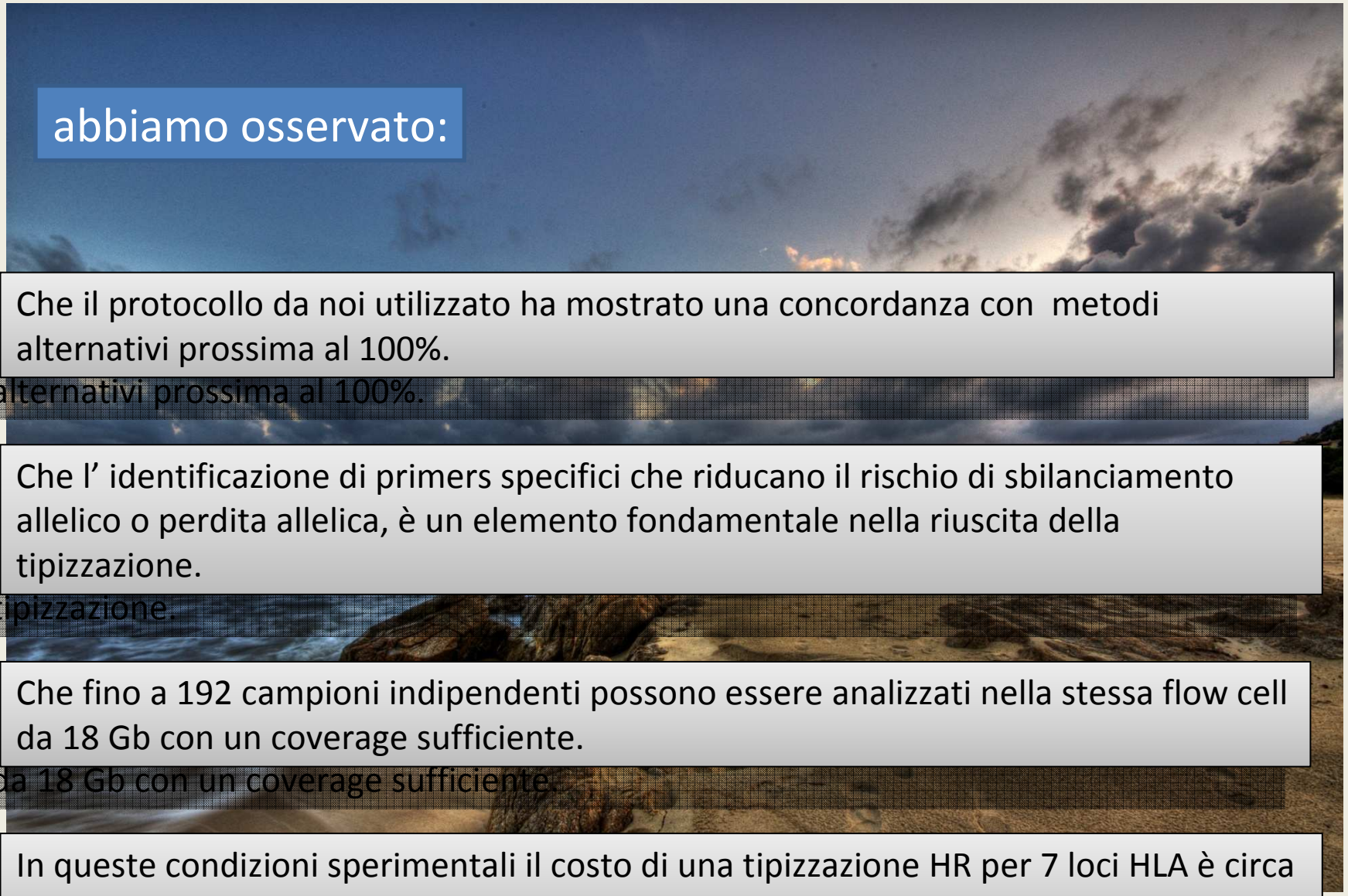
Identificazione di nuovi alleli by Omixon Twin HLA typing

Conclusioni

La NGS può essere usata per lo sviluppo di sistemi di genotipizzazione dei geni HLA producendo molti vantaggi rispetto ai sistemi correnti.

Diventa indispensabile quando è necessario un livello High-throughput come nel caso della tipizzazione per i registri di donatori di midollo osseo o cellule cordonali.

D' altra parte è necessario almeno inizialmente, una attenta programmazione del numero di campioni massimo e minimo da analizzare per flow cell, utilizzo della tagmentazione, size selection dei frammenti, coverage stimato ecc.



abbiamo osservato:

Che il protocollo da noi utilizzato ha mostrato una concordanza con metodi alternativi prossima al 100%.

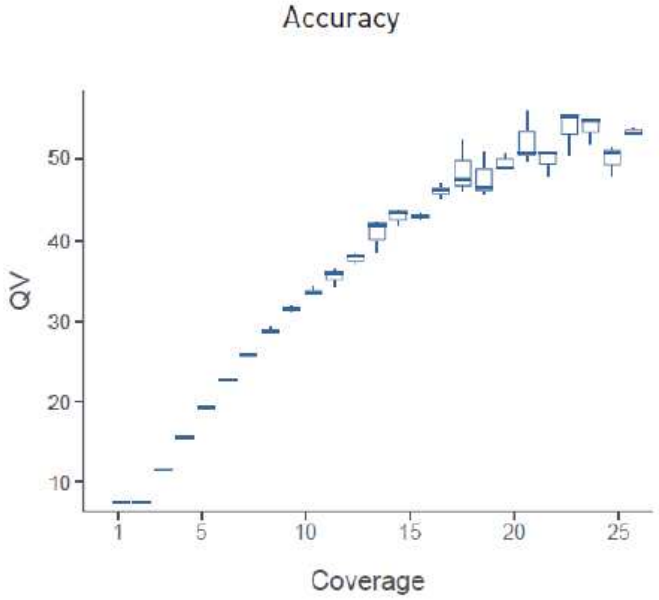
Che l' identificazione di primers specifici che riducano il rischio di sbilanciamento allelico o perdita allelica, è un elemento fondamentale nella riuscita della tipizzazione.

Che fino a 192 campioni indipendenti possono essere analizzati nella stessa flow cell da 18 Gb con un coverage sufficiente.

In queste condizioni sperimentali il costo di una tipizzazione HR per 7 loci HLA è circa 70 euro.

Third Generation Sequencing : Single Molecule Sequencing

Pacbio RS



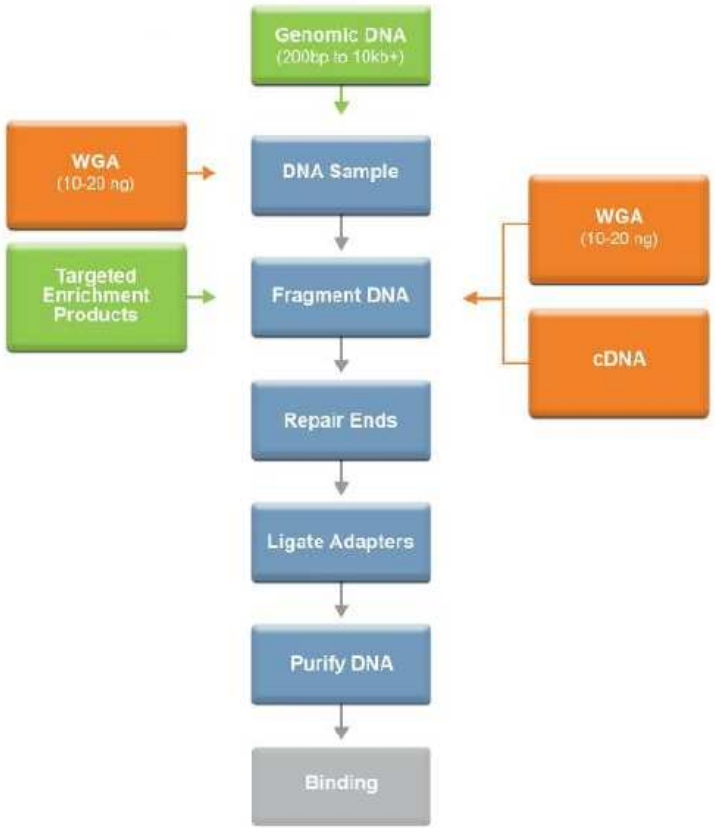
	Pacbio RS
Read Length	3000 - 15,000 bp
Throughput	1 Gb
Reads per run	70,000
Accuracy	95 %
Run Time	30 minutes

Third Generation Sequencing : Single Molecule Sequencing

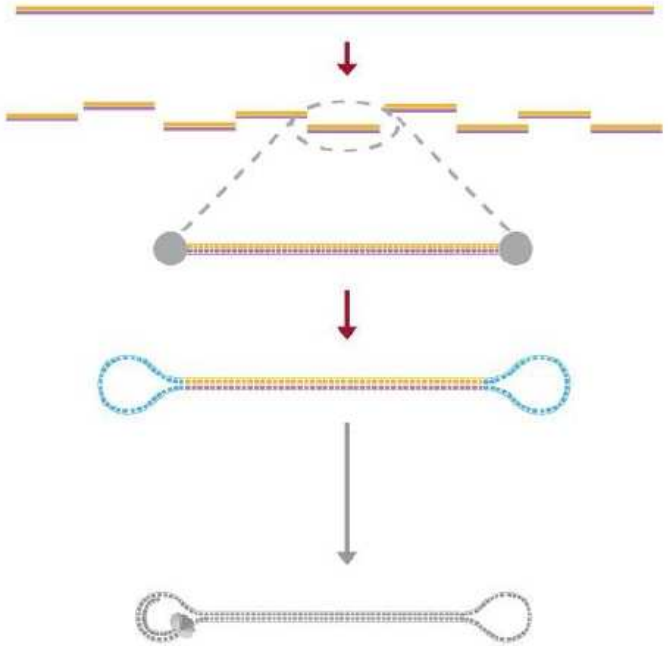
Pacific Biosciences

Workflow : Library preparation → Sequencing

Sample Preparation

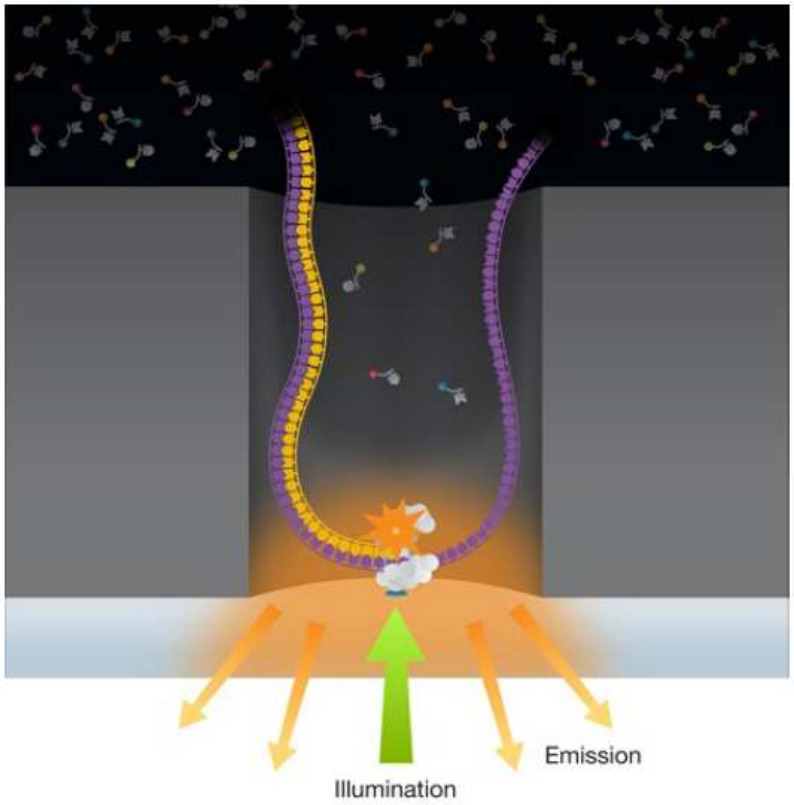


Building of SMRTbell



Third Generation Sequencing : Single Molecule Sequencing

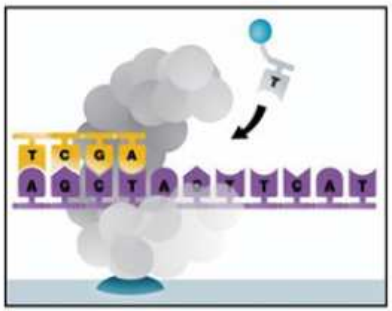
Pacific Biosciences

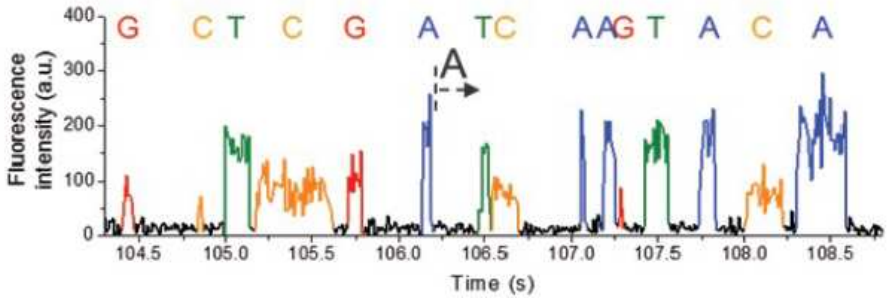


4 nucleotides with different fluorescent dye simultaneous present

2-3 nucleotides/sec
2-3 Kb (up to 50) read length
6 TB data in 30 minutes

laser damages polymerase





Third Generation Sequencing : Single Molecule Sequencing

Oxford Nanopore

Single use cartridge

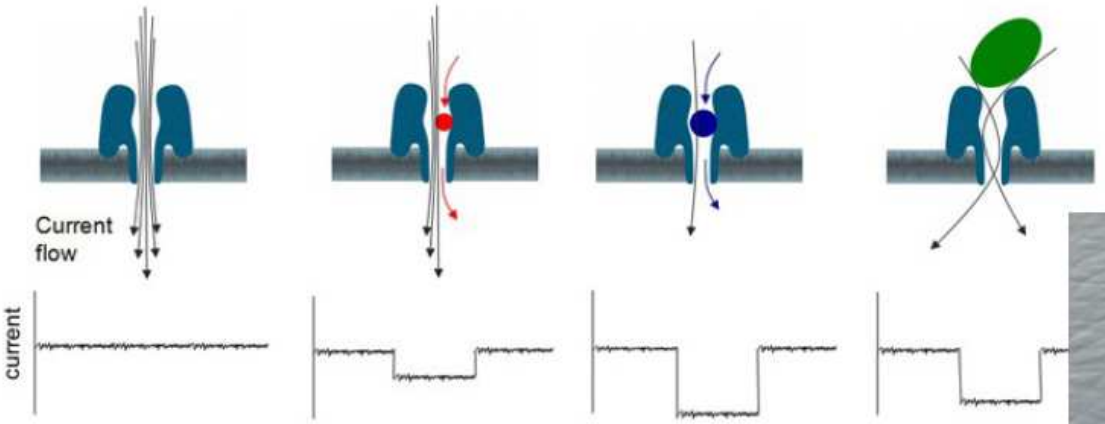
GridION system

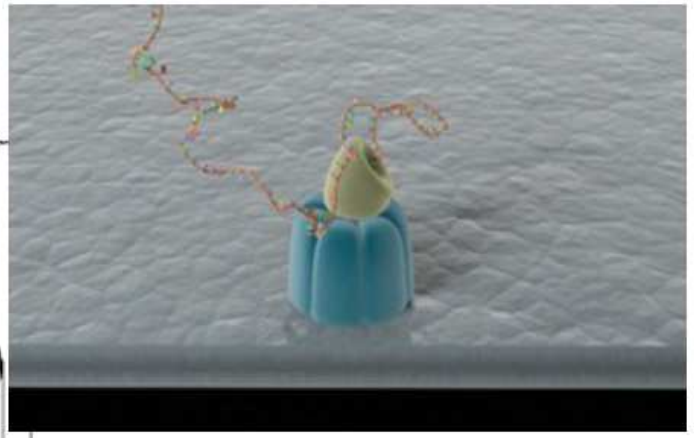
MinION

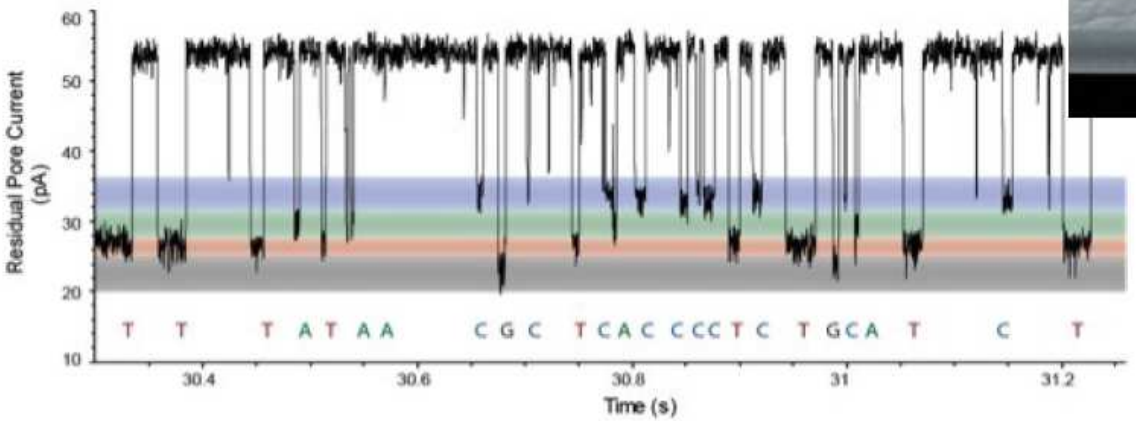
	Nanopore
Read Length	48 kb ?
Throughput	? Gb
Reads per run	2000
Accuracy	75 %
Run Time	? minutes

Third Generation Sequencing : Single Molecule Sequencing

Oxford Nanopore

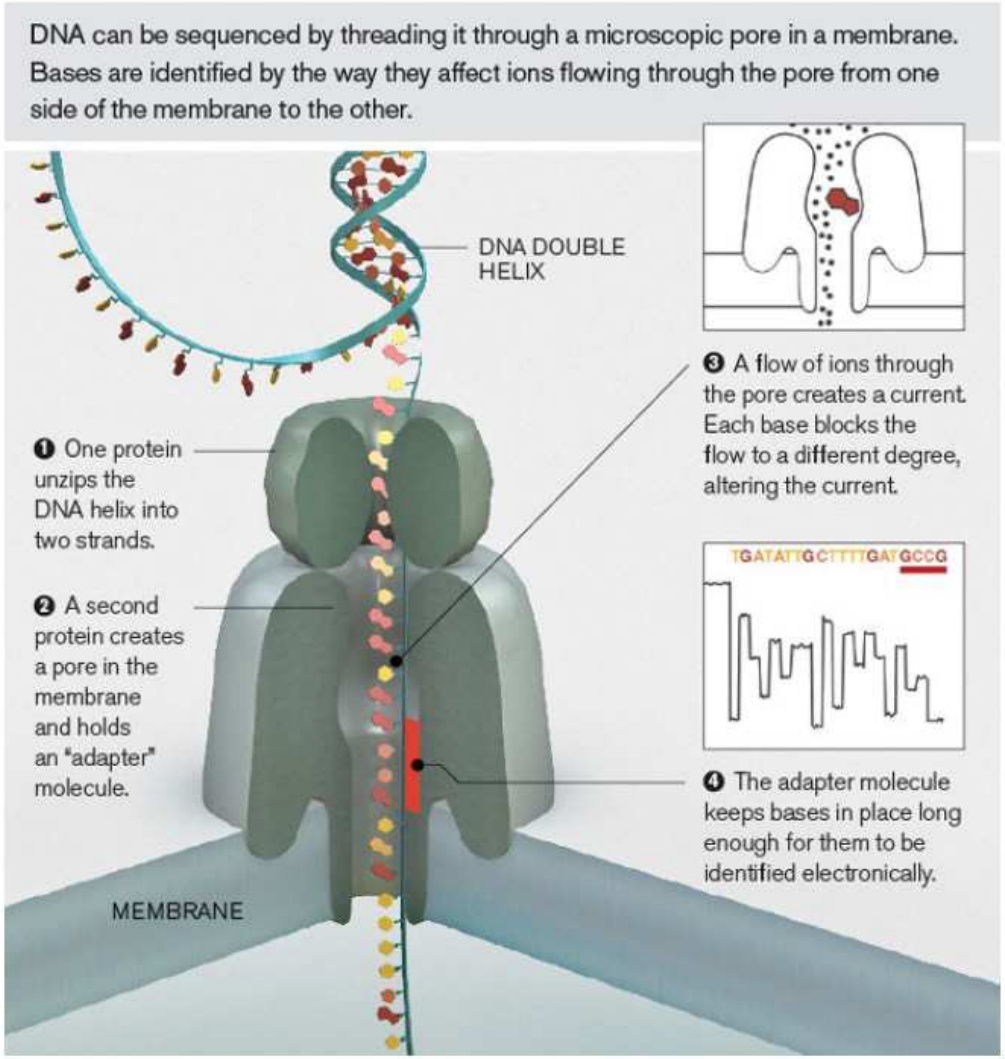






Third Generation Sequencing : Single Molecule Sequencing

Oxford Nanopore



Ringraziamenti

Dr. Sara Lai
Dr. Luisa Cappai
Dr. Francesco Alba
Dr. Federica Cannas
Prof. Carlo Carcassi

Genetica Medica, Università di Cagliari
SC Genetica Medica ASL Cagliari Ospedale Binaghi, 09126 Cagliari

Dr. Annalisa Loizedda
C.N.R. Cagliari

Dr. Roberto Cusano
CRS4, Parco Tecnologico “Polaris”
Pula (CA)