



Istituto Zooprofilattico Sperimentale
del Lazio e della Toscana *M. Aleandri*

Direzione Operativa Diagnostica Generale

*Centro di Riferenza Nazionale per l'Antibioticoresistenza
National Reference Laboratory for Antimicrobial Resistance (Reg. 882/2004/EC)*

Roma, 02/11/2016

Prot. 3267/17 del 23/02/17

All.

Oggetto: Relazione per l'acquisto di apparecchiatura di "Next Generation Sequencing" per applicazioni di "sequenziamento di interi genomi" (incluso Whole Genome Sequencing).

La Struttura Complessa Direzione Operativa Diagnostica Generale, anche in qualità di Centro di Riferenza Nazionale per l'Antibioticoresistenza e National Reference Laboratory for Antimicrobial Resistance (Reg. 2004/882/EC), e Centro di Riferimento Regionale per Agenti zoonosici Speciali, ha la necessità, per gli scopi istituzionali di acquisire uno strumento in grado di eseguire sequenziamenti massivi di genomi di agenti patogeni (batterici, micotici, protozoari) secondo tecnologia definita "Next Generation Sequencing" (NGS).

La presente relazione ha lo scopo di trattare nel dettaglio le seguenti tematiche:

1. Introduzione
2. Motivazioni della Direzione Operativa Diagnostica Generale
3. Logistica
4. Comparazione tecnologia e strumenti sul mercato
 - 4.a. comparazione dei due strumenti Miseq (Illumina) e Ion Torrent S5 (Life Technologies)
 - 4.b. comparazione dei protocolli preliminari al sequenziamento
 - 4.c. analisi dei pro e contro delle caratteristiche tecniche e di accuratezza delle due tecnologie anche in funzione dell'utilizzo prevalente;
5. Strumentazione richiesta, opzioni per piano di acquisto e spesa prevista.
6. Costo reagenti e stima costo di sequenziamento per genoma
7. Possibile utilizzo da parte di altre Direzioni dell'Istituto e possibile evoluzione per scopi istituzionali

1. Introduzione

Lo strumento si rende necessario per eseguire caratterizzazione molecolare per indagini e approfondimenti di epidemiologia molecolare dei principali agenti patogeni e zoonosici e caratterizzare geni ed elementi genetici di antibioticoresistenza correlata.

In particolare trattasi di uno strumento in grado di fornire sequenze di “interi genomi” dei suddetti agenti (p. es. intero patrimonio genetico di un agente patogeno, costituito da milioni di nucleotidi, c. d. “Whole Genome Sequencing”) in modo accurato, rapido ed a costi sostenibili.

Quanto sopra descritto non è realizzabile attraverso tecnologia di sequenziamento tradizionale (“Sanger Sequencing”). E’ necessario inoltre sottolineare che la suddetta tecnologia NGS, non è attualmente disponibile in IZSLT.

Per i volumi attuali della Struttura, sarebbe sufficiente uno strumento con un output medio-piccolo (es. min 10 Gigabytes, con 25 milioni di reads di almeno 100-300 paia di basi teoriche), considerando anche la rapida evoluzione della tecnologia che renderà sempre più accurate e versatili per le varie applicazioni (anche in termini di lunghezza dei frammenti, ovvero “reads”) e che prevederà una “ricambio generazionale“ delle tecnologie e delle relative apparecchiature abbastanza veloce.

2. Motivazioni della Direzione Operativa Diagnostica Generale

Nell’ambito delle attività di sorveglianza e di ricerca recenti e meno recenti, la D. O. Diagnostica Generale già collabora con istituzioni Europee di Sanità Pubblica, Veterinarie e di Food Safety, che già utilizzano estensivamente tale tecnologia. Primo fra tutti, l’European Reference Laboratory for Antimicrobial Resistance (<http://www.crl-ar.eu/>) and Research Group for Genomic Epidemiology (<http://www.food.dtu.dk/english/Research/Genomic-Epidemiology> presso il Center for Genomic Epidemiology, <http://www.genomicepidemiology.org/>, /o Danish Food Institute - Danish Veterinary Institute (DTU-Food).

Obiettivo di questa collaborazione transnazionale è di costituire un network europeo in grado di rispondere con nuovi strumenti alle necessità di sorveglianza epidemiologica e di controllo dei principali agenti patogeni zoonosici, inclusi quelli antibioticoresistenti e multiresistenti.

Allo scopo, la D. O. Diagnostica Generale IZSLT, anche in qualità di NRL-AR, partecipa dall’inizio del 2016, insieme ad altre istituzioni europee del network (coordinate dal DTU-Food, EURL-AR) ad un EFSA Project, “GP/EFSA/AFSCO/2015/01: New approaches in identifying and characterizing microbiological and chemical hazards” (titolo acronimo: ENGAGE).

Focus di tale call nel settore delle malattie trasmissibili all'uomo è di approfondire gli aspetti di epidemiologia molecolare attraverso un uso estensivo di New Generation Sequencing per sequenziare ed assemblare interi genomi dei principali agenti zoonosici, analizzarne le relazioni filogenetiche nei vari settori delle produzioni primarie, degli Alimenti di Origine Animale e nell'Uomo, per scopi di sorveglianza e di controllo a carattere nazionale e comunitario.

<http://www.efsa.europa.eu/it/art36grants/article36/gpefsaafsc0201501>

La proposta del consorzio di Istituzioni che ha presentato il progetto ENGAGE, è stata finanziata con un apposito fondo.

Ulteriori Progetti transnazionali europei nello stesso filone sono quelli del “The Med-Vet European Joint Programme” (EJP) (“One Health – Zoonoses Emerging Threats”, SFS-36-2017), EJP nell’ Horizon 2020 Framework Programme, cui la Struttura ha applicato in due progetti, insieme ad un consorzio di istituzioni Mediche e Veterinarie EU in due differenti topics

(http://www.agence-nationale-recherche.fr/en/anr-funded-project/?tx_lwmsuivibilan_pi2%5BCODE%5D=ANR-16-MRSE-0008)

Inoltre, attraverso l'utilizzo di NGS, la Struttura ha anche già prodotto come leader o come collaboratore varie pubblicazioni su riviste internazionali con adeguato Impact Factor. Si citano di seguito alcuni esempi:

-Franco A, Leekitcharoenphon P, Feltrin F, Alba P, Cordaro G, Iurescia M, Tolli R, D'Incau M, Staffolani M, Di Giannatale E, Hendriksen RS, Battisti

Emergence of a Clonal Lineage of Multidrug-Resistant ESBL-Producing Salmonella Infantis Transmitted from Broilers and Broiler Meat to Humans in Italy between 2011 and 2014.

PloSone 2015 Dec 30;10(12):e0144802. doi: 10.1371/journal.pone.0144802. eCollection 2015.

-Alba P., Feltrin F., Iurescia M., Amoroso R., Donati V., Caprioli A., Leekitcharoenphon P., Hendriksen R., Battisti A. Franco A., 2016. *Transferable colistin resistance mediated by the mcr-1 gene is widespread among Escherichia coli and is emerging in Salmonella in the Italian fattening turkey industry*. ESCMID-ECCMID Congress, Amsterdam, 9-12 April 2016

-Larsen J, Stegger M, Andersen PS, Petersen A, Larsen AR, Westh H, Agersø Y, Fetsch A, Kraushaar B, Käsbohrer A, Feßler AT, Schwarz S, Cuny C, Witte W, Butaye P, Denis O, Haenni M, Madec JY, Jouy E, Laurent F, Battisti A, Franco A, Alba P, Mammina C, Pantosti A, Monaco M, Wagenaar JA, de Boer E, van Duijkeren E, Heck M, Domínguez L, Torres C, Zarazaga M, Price LB, Skov RL. Evidence for Human. *Adaptation and Foodborne Transmission of Livestock-*

Associated Methicillin-Resistant Staphylococcus aureus. Clin Infect Dis. 2016 Nov 15;63(10):1349-1352.

-Alba, F. Feltrin, R. Amoroso, M. Monaco, M.A. Argudín, B. Lauzat, M. Iurescia, L. Sorbara, S. Dottarelli, B. Guerra, A. Pantosti, J. Larsen, A. Battisti, 2016. *Livestock-associated methicillin-resistant Staphylococcus aureus CC398: evidence of zoonotic transmission from livestock to human*. 13th International Conference on Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases, 10-13 May 2016 - Institute of Tropical Medicine, Antwerp, Belgium.

Attualmente, la D. O. Diagnostica Generale usufruisce di un servizio di Next Generation Sequencing “in service” esterno, presso una Società Internazionale “leader nel settore”, che offre sequenziamento attraverso tecnologia NGS basata su varie piattaforme (ognuna più indicata per scopi diversi). Generalmente, tale servizio è tanto più conveniente quanti più genomi vengono sequenziati contemporaneamente (ovvero quanto più efficientemente vengono utilizzati gli output delle apparecchiature, in termini di Gigabytes di dati prodotti in singole corse a seconda della capacità dei chips/flowcells). Di conseguenza, per quanto il servizio sia estremamente valido per invii contemporanei molto numerosi di genomi (es. almeno 70-80 genomi batterici per volta), non lo è per “piccoli progetti” e soprattutto per attività continuative di sorveglianza e ricerca (ad es. 12-24 genomi per volta) che la Struttura è chiamata a fare.

La D. O. Diagnostica Generale, in analogia a quanto già scelto dal network europeo ed internazionale, coordinato dal già citato EURL-AR, Center for Genomic Epidemiology, DK, per l'utilizzo in service della tecnologia NGS sta richiedendo il sequenziamento su piattaforma Illumina, tecnologia leader nel mercato per accuratezza e facilità di utilizzo, considerata attualmente di elezione nella condivisione di dati di sequenze da condividersi in repositories su scala internazionale.

Sia le sequenze NGS con finalità di Whole Genome Sequencing che quelle effettuate per caratterizzazioni di cloni come Multilocus Sequence Typing (solitamente basate su 6-7 geni target “housekeeping”, in funzione di Struttura di popolazione e di relazione tra agenti isolati per scopi di epidemiologia molecolare) o la sua evoluzione come “Core Genome MLST” o “Whole Genome MLST” (basate generalmente su centinaia di geni target) sono sempre più spesso le uniche ammesse per poter depositare in banche dati internazionali nel settore degli agenti di malattia. Un esempio per tutti, la repository per la classificazione degli agenti patogeni enterici, Enterobase <https://enterobase.warwick.ac.uk/>, accetta attualmente sequenze NGS e con sola tecnologia impiegata da Illumina.

L'acquisto dello strumento è pertanto da intendersi come utilizzo in parallelo al servizio esterno (attraverso risorse interne progettuali "dedicate" della Struttura) a seconda delle necessità. Ciò allo scopo di migliorare l'efficienza e nel rispetto dei principi di economicità e di ottimizzazione di risorse finanziarie, umane e strumentali per gli scopi della Struttura e dell'Ente.

Il costo del servizio esterno richiesto NGS sono coperti utilizzando risorse di budget "dedicate" ed interne alla scrivente Struttura.

3. Logistica

Si ritiene che lo strumento debba essere collocato nella D. O. Diagnostica A. L. 114, questo per semplificare le operazioni del personale che per primo utilizzerà lo strumento e perché la stanza in questione può usufruire di una rivalutazione delle caratteristiche dell'impianto di condizionamento (e se necessario di un'implementazione) per il range di temperatura compatibile con le necessità di utilizzo dello stesso (T: 25°C +/- 3°C).

4. Comparazione tecnologia e strumenti sul mercato

Leader in questo settore del mercato, per accuratezza, semplicità e versatilità è la società Illumina (tecnologia *Solexa* and sequencing by synthesis, SBS), che detiene di gran lunga la più importante fetta del mercato internazionale del settore di nostro interesse, e le cui varie tipologie di macchine (specialmente di output medio-piccolo) sono impiegate dalle principali istituzioni di Sanità Pubblica Umana e Veterinaria in Europa e nel Mondo. Tuttavia non è l'unica società attualmente presente in questo settore di mercato, poiché esistono apparecchiature con capacità di Gigabasi simile, basate su tecnologie differenti, di un'altra società, la Life Technologies (Ion Torrent, attualmente è di proprietà della ThermoFisher Scientific).

Strumentazioni con più piccolo output non si ritiene rappresentino un segmento di mercato adatto ai principali scopi istituzionali dell'Ente. (es. MiniSeq Illumina, il cui output massimo arriva solo alla metà di quello del MiSeq Illumina), o progetti Illumina di prossima realizzazione di cui pure è stata data notizia sul mercato, con la stessa tecnologia "Sequence by Synthesis chemistry attualmente usata dalla Società stessa, ma implementata su superficie di "semiconduttore"; si fa riferimento al progetto Firefly Illumina: tale progetto (che forse esiterà in una prima strumentazione solo alla fine del 2017) ha tuttavia un output di solo 1 Gigabasi, ha un target di mercato di uso clinico, ed è pensato soltanto per "targeted sequencing" (determinazioni specifiche di un numero limitato di geni

bersaglio e, es. per oncologia etc). <https://www.genomeweb.com/sequencing-technology/agbt-illumina-provides-additional-details-project-firefly>.

In considerazione del fatto che si ritiene che la scelta debba ricadere su una delle apparecchiature medio-piccole distribuite dalle due società in questione (Illumina e Life Technologies) di seguito si rappresenta:

- 4.a. analisi dei protocolli preliminari (incluse valutazioni su impiego e costi di risorse umane e di consumabili) al sequenziamento vero e proprio del genoma.
- 4.b. un'analisi dei pro e contro delle caratteristiche tecniche e di accuratezza delle due tecnologie e utilizzate dalle due apparecchiature;

4.a. Analisi dei protocolli preliminari (incluse valutazioni su impiego e costi di risorse umane e di consumabili) al sequenziamento vero e proprio del genoma.

Le fasi necessarie alla preparazione del prodotto da sequenziare sono incluse nel così detto “wet lab” e sono di seguito elencate:

- 4.a.1. Estrazione del DNA
- 4.a.2. Preparazione del Tagment DNA o Libraries
- 4.a.3. Amplificazione delle Libraries
- 4.a.4. Purificazione della Libraries
- 4.a.5. Normalizzazione delle Libraries
- 4.a.6. Preparazione Chip di analisi
- 4.a.7. Analisi dei dati

Durante la trattazione delle singole fasi, che segue, verranno illustrati gli step necessari per le operazioni di sequenziamento e verranno messe in evidenza le differenze sostanziali tra i protocolli consigliati dalle due ditte a confronto:

- Kit Nextera XT DNA Library Prep (Sistema ILLUMINA) (Figura 1)
- Ion Xpress™ Plus gDNA Fragment Library Preparation (Sistema Life Technologies) (Figura 2)

La trattazione e la richiesta si basano anche su esperienza diretta del personale della D. O. Diagnostica Generale, che ha seguito un training (esecuzione dei protocolli in laboratorio ed analisi dei dati), durante il twinning presso DTU-Food EURL-AR, nell'ambito della Ricerca Europea ENGAGE, applicando il protocollo Kit Nextera XT DNA Library Prep (Sistema ILLUMINA). Il

sistema Illumina è quello utilizzato dall'EURL-AR (Technical University of Denmark) e dai partner del Consorzio (tra cui Public Health England; Animal and Plant Health Agency, UK; Federal Institute for Risk Assessment, BfR; Germany) e dalla gran parte delle Istituzioni di Sanità Pubblica nel Mondo nel settore d'interesse.

4.a.1. Estrazione del DNA

Per l'esecuzione del protocollo di preparazione si parte ad es. da un microrganismo in purezza (es. isolato batterico su medium Agar), si procede all'ESTRAZIONE DEL DNA mediante KIT di Estrazione (esempio Mini Kit Qiagen o altri), e si esegue una misurazione accurata del DNA con un fluorimetro (es. Qubit). Questo consente di avere la quantità necessaria di DNA da impiegare nei kit di preparazione delle Libraries.

Per le fasi successive è quindi fondamentale partire da una concentrazione nota e definita di DNA.

Se poniamo a confronto i protocolli delle tecnologie ILLUMINA VS Life Technologies, in questa prima fase si osserva che:

1. Applicando il Kit Nextera XT DNA Library Prep (Sistema ILLUMINA)

Sono necessari da 1,8 a 2 ng di DNA totali come riportato sul protocollo del Kit, e vista l'esperienza maturata sappiamo concretamente che si possono applicare diversi metodi di estrazione e purificazione proprio in virtù della bassa quantità di DNA necessario.

2. Applicando il Kit Ion Xpress™ Plus gDNA Fragment Library Preparation (Sistema Life)

Sono necessari fino a 50 ng. E' quindi necessario molto più DNA di partenza e attualmente non ci sono molti Kit di estrazione disponibili sul mercato per alte rese, quindi si avrebbe necessità di materiale di consumo "dedicato".

Inoltre anche qualora fosse possibile utilizzare minori quantità di DNA di quanto raccomandato nei protocolli Ion Xpress, ciò necessiterebbe comunque di accurata "validazione" del nuovo protocollo, (con tempo e costi di consumabili e risorse umane), e dell'introduzione di uno step di amplificazione ulteriore.

4.a.2. Preparazione del Tagment DNA o Libraries

La preparazione del Tagment DNA o Libraries comprende due fasi:

-la frammentazione del DNA

-l'applicazione degli Index Adapter sui frammenti ottenuti (Tagging)

La frammentazione del DNA (Fragment DNA) deve avvenire affinché il DNA possa essere letto in maniera maneggevole deve essere diviso in frammenti di lunghezza contenuta all'interno di un range (generalmente tra 200 e 500bp), quindi diverse copie di DNA vengono frammentate in pezzi di lunghezza diversa di porzioni diverse del genoma.

L'applicazione degli Index Adapter (Tagging DNA) consiste nell'applicazione alle estremità dei frammenti di DNA di sequenze note di nucleotidi considerate come delle Etichette, che consentono il riconoscimento delle porzioni di DNA di quel singolo isolato.

E' possibile fornire alla macchina un insieme definito di isolati in una stessa "flow cell" (corsa di lavoro dello strumento di sequenziamento), ognuno di questi isolati grazie all'applicazione degli Index Adapter verrà univocamente identificato.

Se poniamo a confronto i protocolli delle tecnologie ILLUMINA VS Life Technologies in questa fase si osserva che:

1. Applicando il Kit Nextera XT DNA Library Prep (Sistema ILLUMINA)

La frammentazione e l'applicazione degli Index Adapter avviene in un unico step, ciò comporta minor manipolazione del preparato.

Inoltre il Sistema ILLUMINA permette il sequenziamento di frammenti di dimensioni variabili, comprese tra 150b e 300b, con un alta resa per il completamento del sequenziamento dell'intero genoma (vedi dopo).

2. Applicando il Kit Ion Xpress™ Plus gDNA Fragment Library Preparation (Sistema Life Technologies)

La frammentazione e l'applicazione degli Index Adapter avvengono in due step separati con maggior dispendio di tempo e maggior manipolazione del materiale.

Il Sistema Life Technologies inoltre necessita di una fase di verifica della lunghezza dei frammenti ottenuti attraverso l'impiego di un ulteriore passaggio ("Size selection"); la fase di frammentazione diventa quindi una fase critica da sottoporre a verifica con dispendio di tempo e materiale, a fronte del fatto che per ottimizzare i costi è opportuno/necessario processare NON un unico campione per volta, ma (almeno) 24 campioni di DNA (genomi), talvolta anche di più.

La scelta delle dimensioni dei frammenti utilizzabili, detto appunto "SIZE SELECTION", può essere ottenuta con 3 sistemi differenti, a partire da un sistema meno efficace che si basa sul recupero del DNA da una corsa in gel di agarosio, all'utilizzo di uno strumento dedicato che

attualmente non è presente in Istituto e dovrebbe essere acquistato a parte (es. Bioanalyzer, Agilent).

I sistemi di Size Selection possibili, sono di seguito indicati:

-2.1 E-Gel

-2.2 Pippin Prep

-2.3 Bioanalyzer

2.1 E-Gel: la selezione dei frammenti avviene dopo una corsa su gel di agarosio, si seleziona il peso dei frammenti e si procede al recupero manuale dei frammenti d'interesse. Non è raccomandato per library di frammenti di 500bp. Su un gel di agarosio a doppio pozzetto (sopra e nel mezzo) si carica la library e quando il marker indica che il frammento della lunghezza adeguata/desiderata è arrivato, si recupera. **Gel kit dedicato da acquistarsi a parte** (Figura 3).

2.2 Pippin Prep: è simile al E-Gel ma si ha un recupero automatico dei frammenti d'interesse.

Strumento dedicato non presente in Istituto da acquistarsi a parte (Figura 4).

2.3 Bioanalyzer (Agilent). E' uno strumento che misura le dimensioni dei frammenti di DNA e la loro concentrazione. Con il sistema Life Technologies è consigliato l'utilizzo del Bioanalyzer per analizzare la distribuzione dei frammenti durante diverse fasi di preparazione della library, quali:

-prima della fase di amplificazione delle Library per determinare se necessaria

(alternative Ion Library Equalize Kit -fino 300 b- o qPCR)

-prima della fase di normalizzazione per determinare il fattore di diluizione delle Libraries (vedi dopo).

Quindi per il Sistema Life Technologies, è necessario per lo meno nella fase iniziale della validazione del processo possedere il Bioanalyzer (Strumento dedicato non presente in Istituto da acquistarsi a parte) (Figura 5).

4.a.3. Amplificazione del DNA tagmentato

Affinché la fase di lettura delle sequenze frammentate ed etichettate del DNA del nostro isolato, abbia la resa migliore possibile, verranno prodotte moltissime copie di DNA Etichettato attraverso una fase di amplificazione (PCR).

Se poniamo a confronto i materiali di consumo e i protocolli delle due tecnologie ILLUMINA VS Life Technologies in questa fase si osserva che:

1. Applicando il Kit Nextera XT DNA Library Prep (Sistema ILLUMINA)

L'amplificazione è uno step automatico che non necessita di fasi di verifica.

2. Applicando il Kit Ion Xpress™ Plus gDNA Fragment Library Preparation (Sistema Life Technologies)

L'amplificazione prevede prima la verifica delle dimensioni e della concentrazione dei frammenti attraverso l'applicazione dello strumento **Bioanalyzer** (già descritto) o in alternativa l'utilizzo del **Ion Library Equalize Kit** – (per frammenti fino a fino 300 bp) o l'esecuzione di una **Real-Time PCR quantitativa (qPCR)**.

4.a.4. Purificazione della Library

Con il Sistema ILLUMINA o con il Sistema LIFE Technologies questa fase consente di eliminare i frammenti di DNA troppo piccoli con dei semplici lavaggi con beads magnetici. Non vi sono differenze significative tra i due protocolli.

4.a.5. Normalizzazione delle Library

Questa fase prevede che ogni DNA dei singoli isolati sia presente in eguale concentrazione al fine di ottenere un prodotto con la migliore resa possibile.

Se poniamo a confronto le due tecnologie ILLUMINA VS Life Technologies in questa fase si osserva che:

1. Applicando il Kit Nextera XT DNA Library Prep (Sistema ILLUMINA)

La Normalizzazione avviene per saturazione aggiungendo ai diversi campioni Amp Beads (già utilizzate nella fase di purificazione dei frammenti).

2. Applicando il Kit Ion Xpress™ Plus gDNA Fragment Library Preparation (Sistema Life Technologies)

E' necessario calcolare il fattore di diluizione di ciascuna library, per farlo si può usare il Bioanalyzer (strumento da acquistare a parte, come già ricordato) oppure il qPCR Ion Library TaqMan® Quantitation Kit (Cat. No. 4468802).

4.a.6. Preparazione Chip di analisi

Alla fine del processo, le Libraries così normalizzate possono essere inserite nel Chip di analisi.

Se poniamo a confronto i protocolli delle due tecnologie ILLUMINA VS Life Technologies si può in questa fase notare che:

1. Dopo l'utilizzo del Kit Nextera XT DNA Library Prep (Sistema ILLUMINA)

Si può procedere al caricamento diretto del preparato, da parte del personale tecnico qualificato e adeguatamente addestrato, sul Chip di analisi. Il Chip è così introdotto nel sequenziatore e potrà partire il sequenziamento vero e proprio.

2. Dopo l'utilizzo del Kit Ion Xpress™ Plus gDNA Fragment Library Preparation (Sistema Life Technologies)

Nel Sistema Life Technologies questo procedimento è più complesso, e viene consigliato l'acquisto di uno strumento dedicato per preparazione del Chip. Questo strumento è **Ion Chef, il cui costo si aggira intorno ai 50.000 Euro+IVA, e il cui ingombro è identico a quello del sequenziatore (Figura 6).**

Non è indispensabile, ma l'alternativa è acquistare un'altro strumento denominato "One-touch" meno costoso, **ma molto più indaginoso nell'utilizzo e che richiede perciò un maggior dispendio di risorse umane.**

Conclusioni sulla fase di preparazione delle libraries e caricamento dei chip di analisi

In conclusione, in tutte le fasi susposte (da 4.a.1 a 4.a.6 se si esclude la fase di purificazione delle libraries), l'approccio alla preparazione delle *libraries* e del caricamento dei chip di analisi (anche detti "cartucce") si dimostra più robusto, più rapido e più "user friendly" per l'operatore nel caso dei protocolli Illumina. Inoltre per NGS con tecnologia Life Technologies, Ion Torrent, è necessario investire preliminarmente in ulteriori apparecchiature come **Ion Chef, (e Bioanalyzer Agilent), che fanno lievitare i costi complessivi di almeno 50.000 Euro +IVA.**

4.a.7. Analisi dei dati

Dopo il sequenziamento vero e proprio, il prodotto che ne deriva è un file di testo in formato direttamente in .fastq (Illumina), oppure .bam (Life Technologies), che deve essere trasformato in formato .fastq. Tale file rappresenta ciò che lo strumento ha rilevato nella lettura dei frammenti di DNA tagmentato (libraries), i quali prendono il nome di "reads". Una "read" altro non è che il

CODICE GENETICO di un frammento di DNA del genoma dell'isolato tradotto in CODICE ALFABETICO (Figura 7).

Tipicamente la gran parte delle apparecchiature con tecnologia attualmente "accurata" (ovvero con basso tasso di errore complessivo, sommatoria delle varie componenti di tasso di errore del processo) forniscono un output dell'ordine di milioni di sequenze di "reads" (file di frammenti di DNA) di grandezza variabile, ma comunque nell'ordine di 100-300 basi (in alcuni casi fino a 400-600 basi, ma con tassi di errore già superiori), che poi hanno necessità di essere analizzate ed assemblate per ricostruire le sequenze dell'intero genoma (pangenome), includendo il genoma principale, contenente i geni che conformano il core genome, oltre a geni che fanno parte dell'accessory genome, inclusi quelli su elementi genetici separati come i plasmidi (anche essi tra gli elementi dell'accessory genome).

In Illumina qs fase è diretta, in LIFE l'estensione dei files è .bam, poi va trasformata con apposito software (non è direttamente leggibile come .txt etc.)

4.c. analisi dei pro e contro delle caratteristiche tecniche e di accuratezza delle due tecnologie anche in funzione dell'utilizzo prevalente.

Per fare una valutazione delle caratteristiche tecniche degli strumenti, occorre definire quali siano i parametri da prendere in considerazione.

L'output dello strumento è dato da un numero variabile di sequenze dei frammenti di DNA che, come già detto, prendono il nome di "reads".

L'insieme delle lunghezze di tutte le reads prodotte nella fase di sequenziamento si somma a formare l'output dello strumento, che può variare per il nostro caso da **0,3-15 Gb** (1Gb= 1.000.000.000 basi). L'output dello strumento dipende dalle capacità dello strumento stesso, ma anche dal Chip o cartuccia che viene impiegata che lavora in un range ben definito, e che viene quindi scelta in funzione della quantità di genomi che vogliamo analizzare in una "work flow".

Le reads vengono assemblate in porzioni più grandi dette "contig", che rappresentano parti del genoma di riferimento o parti di un genoma sconosciuto; ogni porzione del genoma completo avrà la possibilità di essere "ricostruito" a partire da molte reads. (Figura 8). Tali "contigs" possono essere quindi assemblati ("genome assembly") per ricostruire l'intero genoma attraverso l'utilizzo di programmi dedicati, che tuttavia hanno bisogno delle necessarie capacità bioinformatiche per essere gestiti.

Il “*coverage*” è il numero medio di volte che ogni read può coprire una parte di genoma, del quale un buon livello sarebbe avere un output di sequenza con almeno 50X di coverage, ovvero con almeno 50 reads che coprono una determinata parte di genoma.

Esempio: genoma batterico da 5Mb ad un coverage teorico di 50x

$5\text{Mb} * 50\text{x di coverage} = 250\text{Mb/campione}$ (1Mb= 1.000.000bp)

Ovvero all'interno della cartuccia ogni campione potrà “occupare” uno “settore” di 250Mb.

Accuratezza della fase di sequenziamento

Nella fase di sequenziamento, lo strumento può commettere errori, e tali errori si manifestano come reads errate, ovvero sequenze che non riproducono correttamente il genoma di riferimento in termini di sequenza di nucleotidi. Gli errori vengono solitamente rappresentati come “**error rates**” (“tassi di errore”) e *l'error rate* è espresso in percentuale. Per le *reads* di dimensioni più piccole (entro le 200-300 basi circa, attualmente) si hanno meno probabilità che nel processo di sequenziamento si vada incontro ad errori.

Di seguito si riportano alcuni lavori scientifici in cui sono stati poste a confronto le diverse tecnologie più utilizzate (semplificando: Società Illumina, tecnologia Solexa detta *Sequence by Synthesis chemistry* con lettura in fluorescenza, e Società Life Technologies, Ion Torrent, tecnologia di sequenziamento tramite “semiconduttori”, con emissione di protoni).

E' necessario rappresentare che la tecnologia utilizzata da Ion Torrent (semiconduttori con rilascio di protoni) non è cambiata nel tempo, tuttavia è cambiata la tipologia di strumento che la utilizza (Ion Torrent PGM, ora Ion Torrent S5/S5XL). La letteratura disponibile arriva al 2014, e al meglio delle nostre ricerche non abbiamo trovato letteratura che confronta il nuovo strumento Ion Torrent S5 con gli altri. La ditta oltre ad aver migliorato la facilità di utilizzo dello strumento (PGM molto indaginoso e con necessità di setting frequente delle condizioni di utilizzo), sostiene che attualmente ha in parte limitato il problema dei tassi di errore proveniente dalla frazione “omopolimeri” (un omopolimero è una macromolecola costituita da un'unica unità di ripetizione (in questo caso, dalla ripetizione dello stesso nucleotide, es. AAAAAA). Tuttavia non abbiamo trovato riscontri oggettivi ed indipendenti di quanto affermato da Life Technologies.

Si riportano dati di letteratura salienti (in cui per Life Technologies, Ion Torrent, è oggetto di confronto lo strumento PGM, che come specificato, utilizza però la stessa tecnologia dell'attuale S5/S5XL):

a. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. Michael A Quail, Miriam Smith, Paul Coupland, Thomas D Otto, Simon R Harris, Thomas R Connor, Anna Bertoni, Harold P Swerdlow and Yong Gu BMC Genomics 2012;13:341 DOI: 10.1186/1471-2164-13-341

“The error rate of the MiSeq is <0.4% vs 1.8% of the Ion Torrent (reference sequences determined by Sanger Sequencing)”

b. Performance comparison of benchtop high-throughput sequencing platforms Nicholas J Loman, Raju V Isra, Timothy J Dallman, Chrystala Constantinidou, Saheer E Gharbia, John Wain & Mark J Pallen Nature Biotechnology 30, 434–439 (2012) doi:10.1038/nbt.2198

“The MiSeq had the highest throughput per run (1.6 Gb/run, 60 Mb/h) and lowest error rates... Unlike the MiSeq, the Ion Torrent PGM and 454 GS Junior both produced homopolymer-associated indel errors (1.5 and 0.38 errors per 100 bases, respectively). ...”

Il tasso di errore molto più contenuto della tecnologia utilizzata da Illumina costituisce un concreto vantaggio allorché l’uso della tecnologia debba comprendere anche “SNP calling” (ovvero l’accurata determinazione delle variazioni di un singolo nucleotide all’interno di una sequenza) per scopi di filogenesi basata su SNP o su Differenze Nucleotidiche (ND) di agenti infettivi. Inoltre è ovviamente un vantaggio ogni volta si debba determinare la sequenza di determinati geni di agenti infettivi, in cui anche mutazioni puntiformi in varie parti del genoma risultano molto rilevanti. Si può citare, come esempio pratico per l’importanza dell’accuratezza complessiva delle sequenze ottenute, la differenza che esiste nella funzione di enzimi (proteine) come beta-lattamasi a spettro esteso, spesso originate da geni mutanti rispetto ad un *wild type*, che modificano in modo sostanziale lo spettro di attività dell’enzima nell’inattivare molecole antibiotiche).

c. Characterizing and measuring bias in sequence data. Michael G Ross, Carsten Russ, Maura Costello, Andrew Hollinger, Niall J Lennon, Ryan Hegarty, Chad Nusbaum and David B Jaffe. Genome Biology 2013;14:R51 DOI: 10.1186/gb-2013-14-5-r51.

I tassi di errore in funzione del contenuto di G+C, specialmente allorché di allontana dal 50% del totale (es. 20-30% o 70-80%, **Figura 9**), della lunghezza degli omopolimeri (**Figura 10**), e **soprattutto il tasso di errore complessivo totale (Figura 11)** sono notevolmente inferiori per la tecnologia Illumina rispetto a Life Technologies Ion Torrent.

Tutti questi lavori indicano per Illumina un error rate complessivo sempre inferiore rispetto a Life Technologies, Ion Torrent. Spesso nell’ordine di almeno 1 logaritmo in base 10.

Infine anche il lavoro più recente “Accuracy of Next Generation Sequencing platforms”, di Fox et al., 2014, <http://www.omicsonline.org/open-access/accuracy-of-next-generation-sequencing-platforms-jngsa.1000106.pdf> è ancora riportato (vedi Figura 12) un log di differenza nella frequenza di errore (tenendo in considerazione il tipo di errore più frequente) tra Illumina e Ion Torrent (10^{-3} vs 10^{-2}).

Conclusioni:

Per le esigenze attuali (segmento di produttività medio-piccolo), il tipo di strumento che la D. O. Diagnostica Generale chiede che venga acquisito è il MiSeq® System della ditta Illumina (così come riportato nel paragrafo 2 della presente relazione).

La richiesta è motivata dal fatto che la tecnologia utilizzata dall'apparecchiatura proposta dalla società Illumina (tecnologia Solexa), è superiore in:

- robustezza e facilità delle fasi di preparazione delle libraries;
- maggior accuratezza della fase di sequenziamento vero e proprio, con minori tassi di errore complessivi (error rates), inclusi e minor tassi di errori indel associati a presenza di omopolimeri, o a contenuto G+C (Guanina + Citosina, ovvero due delle quattro basi azotate di cui è composto il DNA degli organismi viventi), come si evince da letteratura;
- costi complessivi della strumentazione necessaria alle operazioni di sequenziamento che garantiscano l'accuratezza complessiva richiesta.

Inoltre si avrebbe il vantaggio della possibilità di utilizzo della stessa tecnologia Solexa utilizzata da parte del European Reference Laboratory for Antimicrobial Resistance (DTU-Food, DK) e dalle altre istituzioni di Sanità Pubblica del Network del consorzio del progetto di ricerca denominato “ENGAGE”, vincitore della call EFSA GP/EFSA/AFSCO/2015/01 (ad esempio enti governativi come Public Health England, UK; Animal and Plant Health Agency, UK; Federal Institute for Risk Assessment (BfR, Germany). L'acquisizione di un'apparecchiatura che utilizza la stessa tecnologia garantisce una maggiore confrontabilità dei dati prodotti con quelli delle altre istituzioni del consorzio.

Infine, c'è la possibilità di acquisto di un'apparecchiatura ricondizionata e garantita dalla ditta Illumina, che permette di risparmiare circa il 25% sul prezzo di listino, il che è in accordo con i principi di economicità e di efficienza della Pubblica Amministrazione.

5. Strumentazione richiesta, piano di acquisto e spesa prevista.

Per tutto quanto sopra riportato, la tecnologia proposta e commercializzata dalla ditta Illumina rappresenta il gold standard attuale nel settore di utilizzo (NGS/WGS di agente infettivi).

Pertanto, per le esigenze di produttività attuali, la D. O. Diagnostica Generale chiede che venga acquisito è il MiSeq® System della ditta Illumina, ricondizionato.

L'utilizzo di tale strumento è da intendersi per attività continuative di sorveglianza e ricerca che la Struttura è chiamata a fare in ambito nazionale ed Europeo, data necessità di processare con maggior frequenza un numero più limitato di genomi (ad es. 12-24 genomi per volta) da sequenziare rispetto all'ottimizzazione del servizio esterno di cui già si serve (e per il quale sono necessari un minimo di 80-90 genomi per volta, ovvero la capacità di una "lane" di una macchina a output elevato come ad es. Hiseq2000, allo scopo di contenere i costi del servizio).

Tale sistema è molto flessibile, ed oltre al sequenziamento completo di piccoli genomi (es. agenti infettivi come virus, batteri, miceti, protozoi etc.) può eseguire sequenze di ampliconi, sequenze di geni target specifici, possiede:

- un output range di 0,3-15 Gb (a seconda delle cartucce disponibili)
- un tempo di esecuzione della corsa (flow cell) che varia dalle 5 alle 55 ore
- numero di reads per flow cell fino a 25 milioni
- lunghezza delle reads fino a 2x300bp.

Le diverse opzioni di acquisizione possibili sono (come da colloqui preliminari con il distributore):

-Acquisto strumento nuovo:

Il costo Illumina per MiSeq è di 98.000 Euro +IVA con 12 mesi di garanzia "full risk" inclusa, più altri 12 mesi "full risk" al costo di 13000 Euro +IVA.

L'assistenza territoriale Illumina attualmente dispone di 2 specialist su territorio dell'Italia Centrale e supporto telefonico full time con intervento sullo strumento da remoto.

Disponibilità specialist per "wet lab" (preparazione campioni vedi paragrafo 7) e "dry lab"(per es. possibilità di utilizzo di Cloud di Illumina che procede alla processazione dei dati fino alla produzione di files Fastq, e fino a 1Tb di capacità.

-Acquisto strumento ricondizionato del parco Illumina:

Il costo Illumina per MiSeq è di 75.000 Euro +IVA con 12 mesi di garanzia "full risk" inclusa, più altri 12 mesi "full risk" al costo di 13000 Euro +IVA (che IZSLT solitamente chiede per le apparecchiature di alta tecnologia).

L'assistenza territoriale Illumina attualmente dispone di 2 specialist su territorio dell'Italia Centrale e supporto telefonico full time con intervento sullo strumento da remoto.

Disponibilità specialist per "wet lab" (preparazione campioni vedi paragrafo 7), "dry lab" e "data analysis & storage" (per es. possibilità di utilizzo di Cloud di Illumina che procede alla processazione dei dati fino alla produzione di files Fastq, e fino a 1Tb di capacità).

-Noleggio (con possibilità di riscatto):

Spesa stimata 4500-5000 Euro +IVA mensile (variabile in relazione al mercato dei tassi di interesse al momento della transazione). In accordo con la Direzione Aziendale, si potrà decidere, ad esempio, di scegliere un noleggio per 36 mesi (oppure inferiore, es. 12, 24, ma economicamente meno conveniente), che già include garanzia "full risk" annuale ai costi succitati.

Nel caso si volesse optare per un noleggio di 36 mesi, i costi di noleggio dell'apparecchio nuovo, sarebbero circa 4500+IVA mensili per un totale di: $4500 \times 1,22 \times 36 = 197.640$ Euro (vs 151.283 Euro calcolato su acquisto strumento nuovo + IVA+ 24 mesi aggiuntivi di garanzia full risk). Il riscatto finale sarebbe di circa 1% del costo dello strumento.

I costi di noleggio dell'apparecchio ricondizionato per 36 mesi si aggirerebbero intorno a 4000 Euro mensili+IVA, per un totale di $4000 \times 1,22 \times 36 = 175.600$ Euro (vs 151.283 Euro calcolato su acquisto strumento nuovo + IVA+ 24 mesi aggiuntivi di garanzia full risk).

Per le modalità di assistenza territoriale e disponibilità di specialists e di spazio su Cloud, vale quanto sopra già detto.

[Nota: A puro titolo di confronto dei costi: di listino, Ion Torrent S5 costa 80.000 Euro+IVA; S5XL 130.000+IVA (maggiore capacità di hard disk e RAM più elevata, con velocità di output dati, da preferirsi, Ion CHEF 50.000 Euro+ IVA): Totale setting da preferirsi: **180.000 Euro+IVA vs MiSeq nuovo 98.000 Euro+IVA**].

Per le varie opzioni di acquisizione di MiSeq Illumina, la Struttura mette a disposizione fondi su proprio budget (es. noleggio attraverso fondi dedicati apparecchiature su DIA/DIG/LT0715 (Ricerca Corrente 2015); DIA/DIG/8AMR2 (Progetto Ministero Salute Monitoraggio antibioticoresistenza).

6. Costo reagenti e stima di costo di sequenziamento per genoma

Quanto costa (consumabili + personale) fare WGS di un genoma di medium size (5Mb, es Escherichia coli, Salmonella; S. aureus ha un genoma un po' più piccolo, quindi l'efficienza di una cartuccia/flow cell è un po' superiore).

Per preparare una libreria di DNA con Nextera XT, (target size, range: 300bp fino 200Mb), il costo è di €32,00/campione (tutti i costi sono IVA esclusa).

Per correre le librerie di genomi, su Miseq si può scegliere ben 6 cartucce diverse che vanno **da 300Mbasi** (costo per corsa: € 275) **a 15Gbasi** (costo per corsa € 1470) di produttività.

Esempio: genoma batterico da 5Mbasi ad un coverage teorico di 50x

costo libreria NexteraXT: € 32,00/campione

costo corsa: 5Mb * 50x di coverage = 250Mb\campione

Range di costo\campione: Max €275,00\campione Min: €24,50/campione. Il costo, come già chiarito in precedenza nella relazione, varia a seconda della cartuccia e dei campioni che vengono caricati: a titolo di esempio, si riportano i costi delle cartucce più performanti:

MS-103-1002 (1,2Gb – 4 M reads) corre massimo 5 campioni costo cartuccia € 417,00

MS-102-2002 (4.5Gb – 15 M reads) corre massimo 18 campioni costo cartuccia € 996,00

MS-102-2003 (8,5Gb – 15 M reads) corre massimo 35 campioni costo cartuccia € 1.114,00

MS-102-3003 (15Gb – 25 M reads) corre massimo 60 campioni costo cartuccia € 1.470,00

Tuttavia, per un High Coverage (es. Standard di Qualità Consortium ENGAGE) e per un'accuratezza che consenta Qualità Control values di consenso Europeo ed Internazionale:

MS-102-3003 (15Gb – 25 M reads): 36 campioni (5Mb)

Il costo complessivo di un genoma da 5Mb pertanto è stimato in:

€ 40,83 + library prep € 32,00= € 72,83.

Oltre ai costi dei consumabili, si stima che per la preparazione del DNA e delle libraries, e per la corsa della flow cell nell'apparecchiatura i costi di personale (Tecnico Lab. Cat D) sono:

0,5 h per genoma.

Ovvero circa 9 Euro di costo di personale per genoma =0,5 h*18,00 Euro/h, oneri riflessi inclusi).

7. Possibile utilizzo da parte di altre Direzioni dell'Istituto e possibile evoluzione per scopi istituzionali

Si sottolinea che dell'acquisizione dell'apparecchiatura da parte della D. O. Diagnostica Generale, potrebbero fruirne anche altre Strutture Complesse per scopi istituzionali e di ricerca in altri settori, inclusi i possibili sviluppi futuri in settori non direttamente riconducibili alle malattie infettive, allorché vi sia evidenza che vi sia richiesta di mercato ed i costi decrescano. Ciò è possibile grazie alla particolare flessibilità dello strumento ad operare sequenziamenti di genomi medio-piccoli, o per evidenziare sequenze target di interesse (*targeted sequencing*) anche di organismi superiori.

Inizialmente, tuttavia, visti i costi complessivi attuali, si ritiene che le finalità di utilizzo della tecnologia (e dell'apparecchiatura) debbano essere quelle di sorveglianza e ricerca su base nazionale. Pertanto, naturali utilizzatori, oltre la D. O. Diagnostica Generale, appaiono essere Strutture Complesse quali la D. O. Diagnostica Malattie Virali (CRN Malattie degli Equidi) e US Biotecnologie e Virologia (CRN e NRL per la ricerca di OGM).

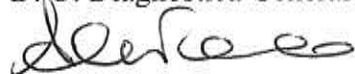
La D. O. Diagnostica Generale, una volta acquisita l'apparecchiatura si rende disponibile, in collaborazione con US Biotecnologie, ad organizzare un momento formativo per presentare la tecnologia e le sue più pertinenti applicazioni in Sanità Animale e Sanità Pubblica Veterinaria al personale dell'Istituto.

Inoltre, la D. O. Diagnostica Generale, una volta portati a termine gli obiettivi che la vedono impegnata nelle applicazioni progettuali in corso basate su tale tecnologia, si rende disponibile all'addestramento di personale dei Centri di Referenza succitati per la preparazione del DNA (frammentazione, tagging, caricamento e processazione del DNA per il sequenziamento) allo scopo di favorire l'acquisizione delle competenze tecniche per le operazioni preliminari al sequenziamento. Come ulteriore evoluzione a medio termine, ed in accordo con la Direzione Aziendale, potrà successivamente condividere la tecnologia NGS con altre Strutture dell'Ente, nelle suddette applicazioni di SPV e per scopi di filiera a carattere (bi)-regionale.

In fede

Alessia Franco, DVM

D. O. Diagnostica Generale

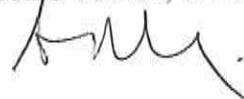


In fede

Il Responsabile

D. O. Diagnostica Generale, CRN-AR, NRL-AR

Antonio Battisti, DVM



Di seguito, le figure citate nel testo

Figura 1 Kit Nextera XT DNA Library Prep (Sistema ILLUMINA)

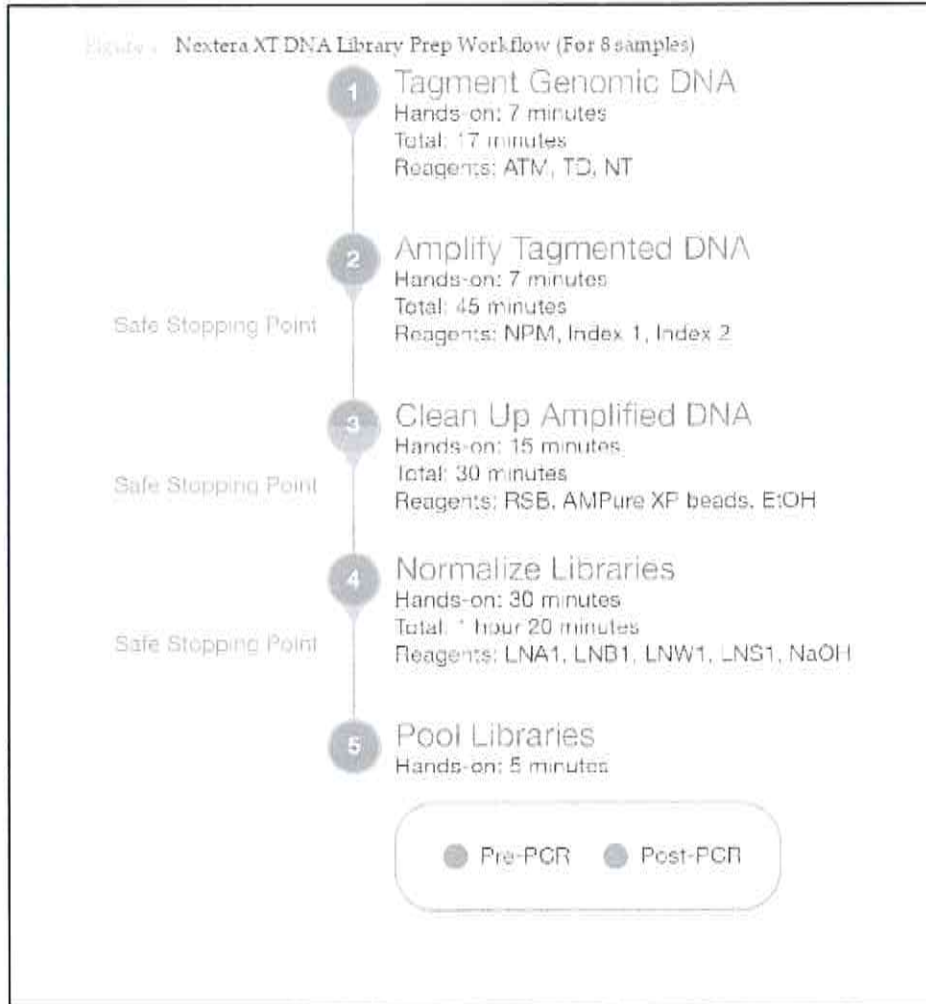


Figura 2 Kit Ion Xpress™ Plus gDNA Fragment Library Preparation (Sistema Life Technologies)

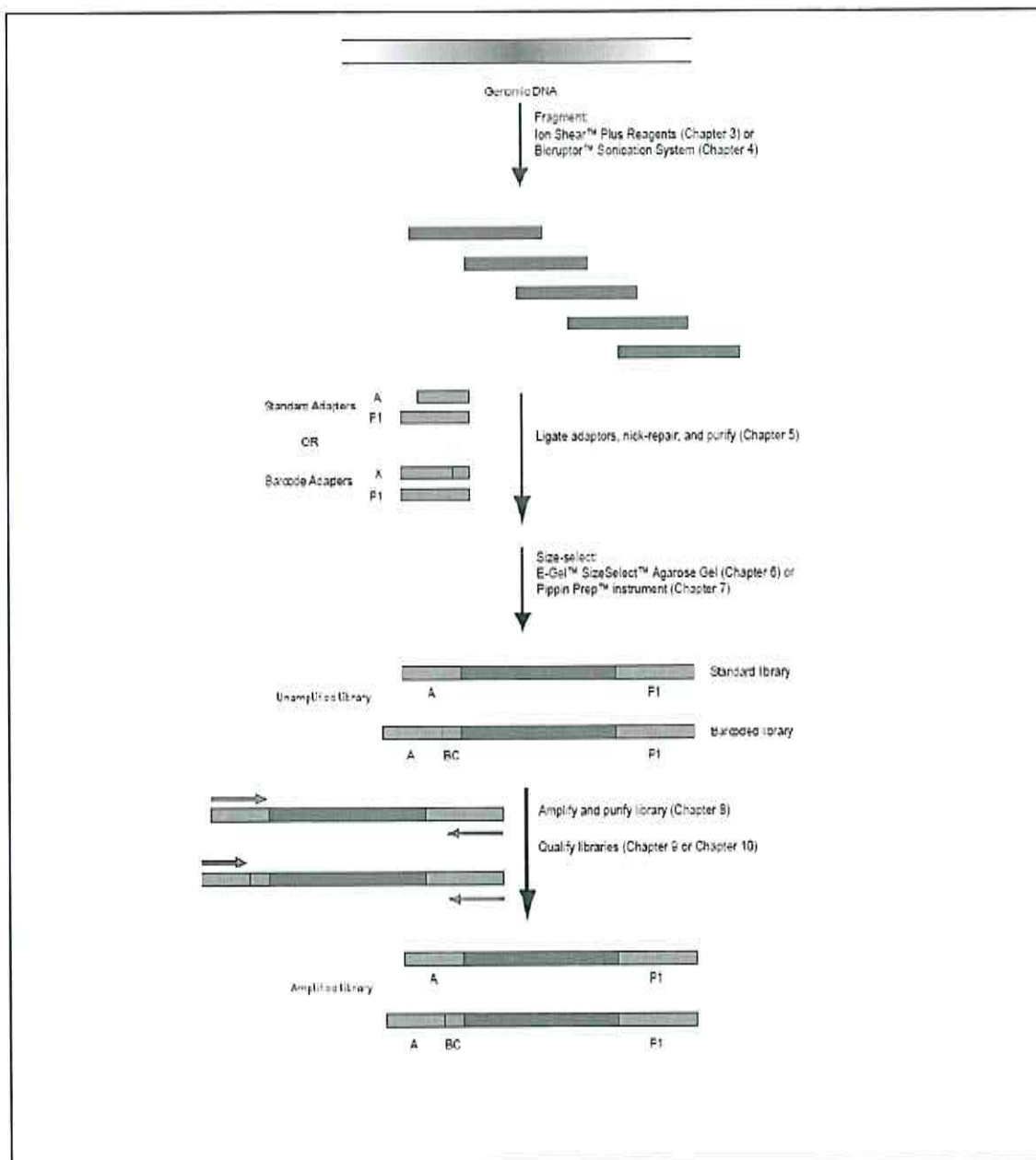


Figura 3 E-Gel

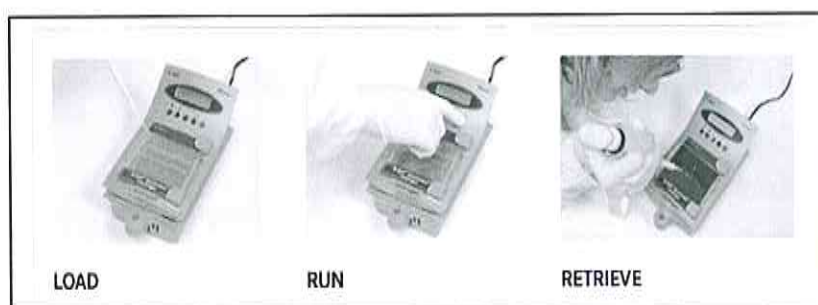


Figura 4 Pippin Prep



Figura 5 Bioanalyzer



Figura 6 Ion Chef e Ion Torrent S5

<p>Dual Function: Ion AmpliSeq Library (15 mins) Template Preparation (15 mins)</p>	<p>Sequencing</p>
<p>30 min hands-on time</p>	<p>15 min hands-on time</p>

Figura 7 Esempio di file .fastq

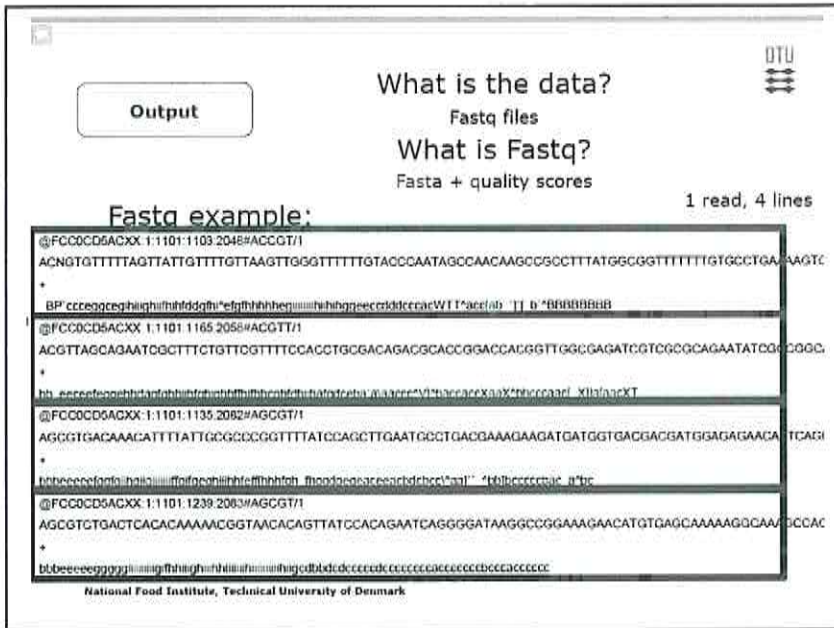


Figura 8. Esempio di come le “reads” (segmenti blu) dopo assemblaggio vadano a formare frazioni di genoma più grande “contigs” (segmenti verdi) e per finire vengano messi a confronto con un genoma di riferimento (linea continua rossa)

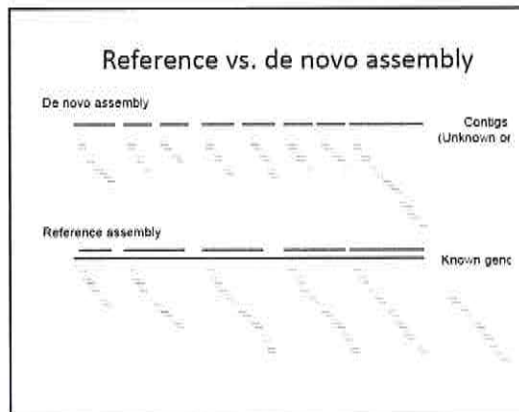
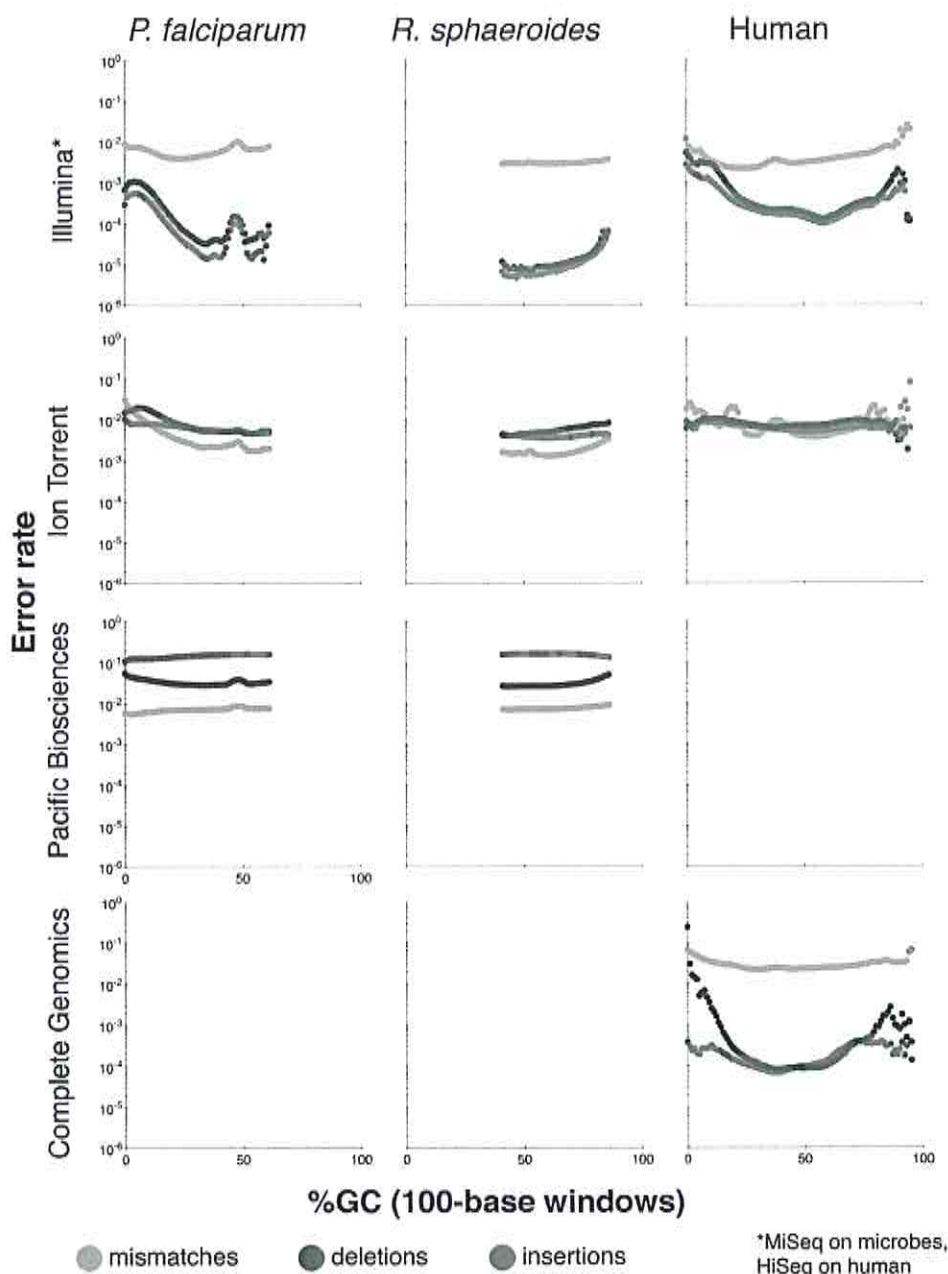


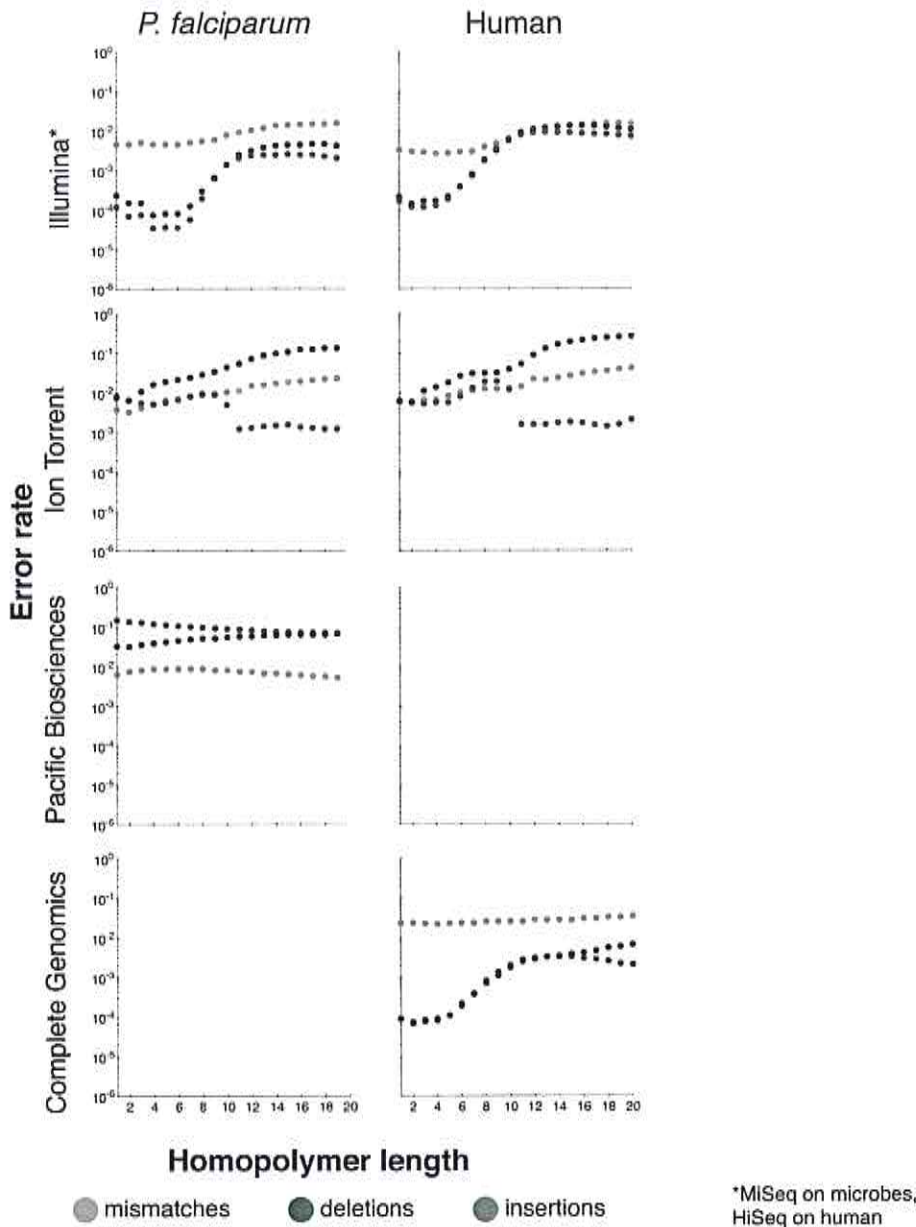
Figura 9. Da “Characterizing and measuring bias in sequence data” (Ross et al., 2013): <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-5-r51>



Legenda

[Figure 4. Error rates as a function of GC composition. Each graph shows mismatch (light blue), deletion (dark blue), and insertion (maroon) rates (y-axis) as a function of GC composition (x-axis). Data are shown for the Ion Torrent PGM from three organisms (*P. falciparum*, *R. sphaeroides*, and human), for the Illumina MiSeq on the two microbes, for the Illumina HiSeq on human, for Pacific Biosciences from the two microbes and from Complete Genomics for human (Table 2, data sets 1 to 3, 7 to 9, and 14 to 16). For human we note that bona fide differences between the sample and the reference sequence were recorded as errors. Error rates are only plotted for GC percentages for which there are at least 1,000 100-base windows in the genome.]

Figura 10: Da “Characterizing and measuring bias in sequence data” (da Ross et al., 2013): <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-5-r51>



Legenda:

[Figure 5. Error rates as a function of homopolymer length. Each graph shows mismatch (light blue), deletion (dark blue), and insertion (maroon) rates (y-axis) within homopolymers of various lengths (x-axis). Data are plotted from *P. falciparum* and human as available (Table 2, data sets 1 to 3 and 14 to 16). For human we note that bona fide differences between the sample and the reference sequence were recorded as errors.]

Figura 11. Da “Characterizing and measuring bias in sequence data” (da Ross et al., 2013):
<http://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-5-r51>

Data set			Fractional error rate			
Sample	#	Platform	Mismatches	Deletions	Insertions	Total
P. falciparum	1	Illumina MiSeq	0.0046	0.00021	0.00011	0.0049
	2	Ion Torrent PGM	0.0038	0.0090	0.0068	0.020
	3	Pacific Biosciences RS	0.0068	0.033	0.14	0.18
E. coli	4	Illumina MiSeq	0.0036	0.0000097	0.0000051	0.0037
	5	Ion Torrent PGM	0.0018	0.0053	0.0044	0.012
	6	Pacific Biosciences RS	0.0077	0.032	0.17	0.21
R. sphaeroides	7	Illumina MiSeq	0.0030	0.000018	0.0000089	0.0030
	8	Ion Torrent PGM	0.0014	0.0055	0.0037	0.011
	9	Pacific Biosciences RS	0.0076	0.029	0.16	0.20
Human	14	Illumina HiSeq	0.0030	0.00023	0.00017	0.0034
	15	Ion Torrent PGM	0.0060	0.0069	0.0057	0.019
	16	Complete Genomics	0.023	0.000099	0.000091	0.024

In Table 5, we show the fractional rates of mismatch, deletion, and insertion, computed relative to coverage, inferred by comparison to the reference sequences. For human we note that *bona fide* differences between the sample and reference sequence were recorded as errors.

Figura 12. Da “Accuracy of Next Generation Sequencing platforms”, Fox et al., 2014
<http://www.omicsonline.org/open-access/accuracy-of-next-generation-sequencing-platforms-jngsa.1000106.pdf>

Commercial Platform	Most Frequent Error Type	Error Frequency
Capillary sequencing	single nucleotide substitutions	10 ⁻¹
454 GS Junior	Deletions	10 ⁻²
PacBio RS	CG deletions	10 ⁻²
Ion Torrent PGM	Short deletions	10 ⁻²
Solid	A-T bias	2*10 ⁻²
Illumina MiSeq	single nucleotide substitutions	10 ⁻³
Illumina HiSeq2000	single nucleotide substitutions	10 ⁻³