Genomeeting 2016 Análisis de datos RNA-Seq



Breviario sobre tecnologías de Secuenciación Masiva

M. en C. Israel Aguilar Ordóñez (<u>iaguilar@wintergenomics.com</u>)

Instituto Nacional de Medicina Genómica Octubre 2016

También conocido como...

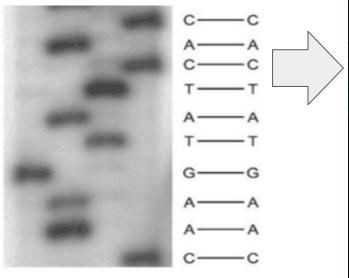
- Secuenciación Masiva

- Massive Parallel Sequencing

- Next Generation Sequencing (NGS)

Secuenciación

¿Cómo pasamos de esto...?



¿A esto?

```
40421551 40421561 40421571 40421581 40421591 40421601 40421611 40421621 40421631 40421641 40421651 40421651 40421671 40421671 40421681 40421691 40421791 40421711 4042
721tttgagcagacctatataagatggttatgaagattcacaccagcggctcatgcctgtgatcccagcactttgggaggctgaggcaagtggagcacctgagatcatgagttcaagaccagcctggccaacatggtgaaaccccatctctactaaagatacaaaaattatccaggtgtggtg
                                                                                                                     cagectggccaacatggtgaaaccccatctctactaaaga ACAAAAATTATCCAGGTGTGG
          GACCTATATANGAT GGTTATGANGAT CACACAG TGGCTC CC TG TGATCCCAGGACTTTGGGAGGC TGAGGCANG TGGAG ACC TGAGATCATGAGTTCANGACCAGC TGGACCAACT TGGAGACAAAAATTATCCAGGTG
     GANCAG ATATAMGA TGGTTA GAAGATTCACACAG TGGCCCATGCC tgateccageacttigggagg TGAGGCAAG TGGAGCCCCTGAGTCATGAGCCA GCCAACATGGTGAAACCCCATCCTTACTAAAGATACAAAA ATCCAGGTGTG
                CAGA TGGT TA TGAAGAT TCACACAG TGGC TCA TGCC TGT A TCCCAGCACT
     agcctggccaacatggtgaaaccccatctctactaaagat AAAATTATCCAGGTGTGG
                                                                                                                        GCCTGGCCAACATGGTGAAA CCCATCTCTACTAAAGATACAAAAA
CCTGGCCAACATGGTGAAACCCCATCTCTACTAAAGATAC
                                                                                                                              CCAACATGGTGAAACCCCCATCTCTACTAAAGATACAAAAA
                                          CAGTGGCTCATGCCTGTGAT
                                                                                                                               CAACATGGTGAAACCCCATCTCTACTAAAGATACAAAAAT
     GAACAGACCTATATAAGATGGTTATGAAGA
                                               GCTCATGCCTGTTATC CTTTGGGGGGCTGAGGCANGTGGAGACCTGAGATCATGA
ACA
CTCTTGCCTGTGATCCCAGCACTTTGGGAGGCTGAGGCANTGTGGAGACCCTGAGATCATGATCAAGACCAGCCTGGCCA
CTCATGCCTGATCCCAGCACTTTGGGAGGCTGAGGCAATTGGAGCACCTGAGATCATGAGTCAAGACCAGCCTGGCCA
     GAACAGACCTATCTAAGATGGTTATGAAGA
     GAACAGACCTA TATAAGA TGGTTA TGAAGA
                                                                                                                                    TGGTGAAACCCCATCTCTACTAAAGATACAAAAATTATCC
                                                        GTGATCCCAGCACTTTGGGAGGCTGAGGCAAGTGGAGCAC
GATCATGAGTTCAAGACCCGCCTGGCCAACATGGTGAAAC ccatctctactaaagatacaaaaattatccaggtc
                                                                                                                                  CATGGTGAAACCCCATCTCT CTAAAGATACAAAAATTATCCAGGTGTGG
                                                                                                                                       GTGAAACCCCATCTCTACTAAAGATACAAAAATTATCCAG
                                                                                                                                      gtgaaaccccatctctactaaagatacaaaaattatccag
                                                                                                                                       gtgaaaccgtgtctctac aaagatactaaaattatccaggtgtg
GAAATCCCATCTCTACTAAAGATACAAAAATTATCCAGGT
                                                                                                                                        GAAACCCCATCTCTACTAAAGATACAAAAATTATCCAGG
                                                                                                                                        GAAACCCCATCTCTACTAAATAAACA
                                                                                                                                           AACCCCATCTCTACTAAAGATACAAAAATTATCCAGGTGT
                                                                                                                                          AATCCCATCTCTACTAAATATACAAAAATTATCCAGGTG
                                                                                                                                           AACCCCATCTCTACTAAAGATACAAAAATTATCCAGG
                                                                                                                                           ACCCCGTTTCTACTAAAGATACAAAAATTATCCAGGTGTG
                                                                                                                                           accccatctctactaaagatacaaaaattatccaggtgtg
                                                                                                                                               CATC CTAA AAAGA ACAAAAATTA CCAGG G
                                                                                                                                               CGTCTCTACTAAAGATACAAAAATTATCCAGGTGTGG
                                                                                                                                               CATCTCTACTAAAGATACAAAAATTATCCAGGTG
```

Compañías que ofrecen instrumentos NGS

Las diferencias radican en:

- Creación de bibliotecas
- Química de la secuenciación
- Métodos de detección
- Cantidad y calidad de datos que arroja el instrumento









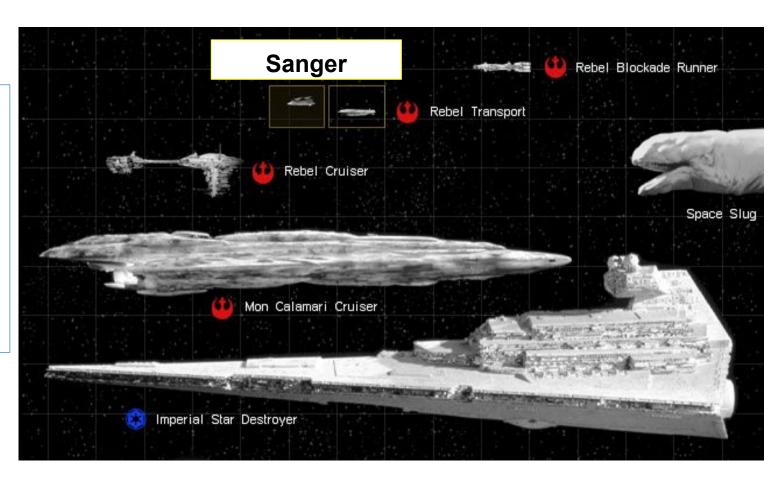




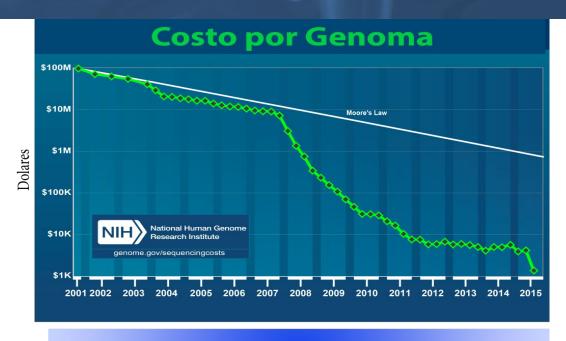
Noblegen

Las tecnologías NGS son un gran paso en la exploración genómica





Impacto de las tecnologías NGS





Primera Generación

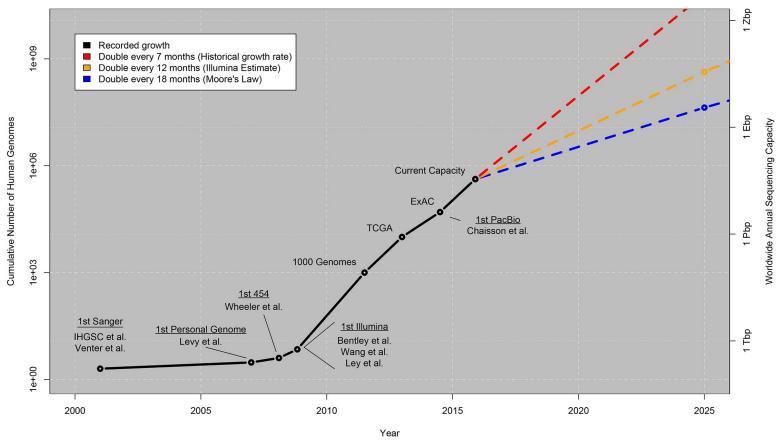
Segunda Generación

Tercera Generación Sanger

454 SOLiD Ion Torrent
Illumina Complete Genomics

Pacific Biosciences
Nanopore

Cantidad de datos genómicos a nivel mundial

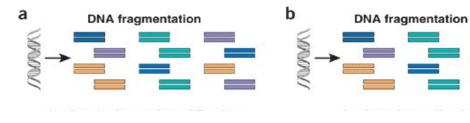


Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., ... & Robinson, G. E. (2015). Big data: astronomical or genomical? PLoS Biol, 13(7), e1002195.



Sanger

NGS (illumina)

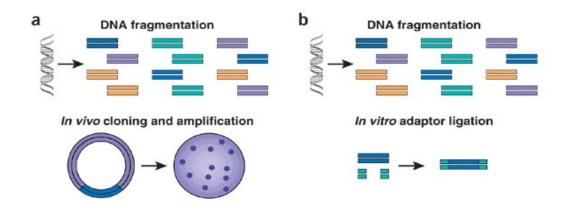


Principales diferencias entre secuenciación Sanger y NGS

Obtención y fragmentación de la muestra.

Sanger

NGS (illumina)



Etiquetado y separación de fragmentos

¿Qué es un adaptador?



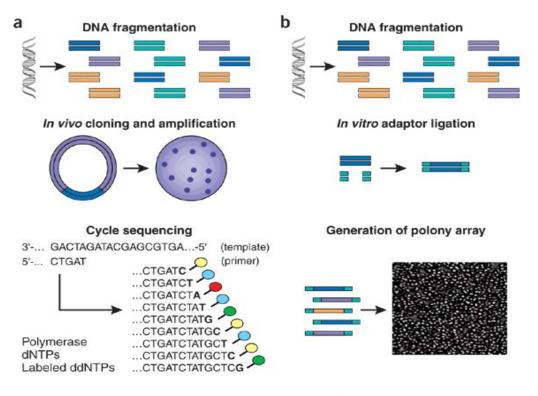
(illumina) **DNA fragmentation DNA fragmentation** In vivo cloning and amplification In vitro adaptor ligation Cycle sequencing Generation of polony array 3'-... GACTAGATACGAGCGTGA...-5' (template) 5'-... CTGAT (primer) Polymerase dNTPs Labeled ddNTPs

Sanger

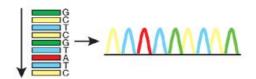
NGS

Principales diferencias entre secuenciación Sanger y NGS

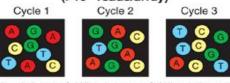
Generación de señal



Electrophorsesis (1 read/capillary)



Cyclic array sequencing (>10⁶ reads/array)



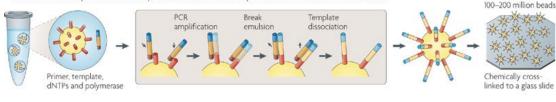
What is base 1? What is base 2? What is base 3?

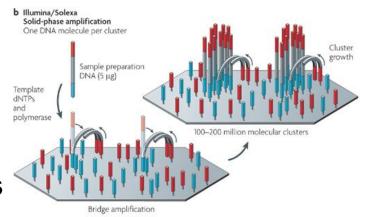
Detección de la señal

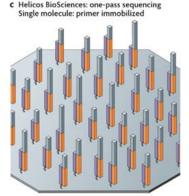
Diferencias entre NGS

Dependiendo la tecnología, existen diferentes estrategias para inmovilizar y secuenciar los fragmentos de DNA

a Roche/454, Life/APG, Polonator
 Emulsion PCR
 One DNA molecule per bead. Clonal amplification to thousands of copies occurs in microreactors in an emulsion



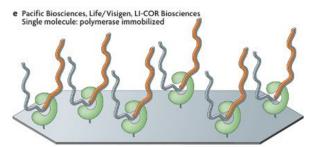




Billions of primed, single-molecule templates

d Helicos BioSciences: two-pass sequencing Single molecule: template immobilized

Billions of primed, single-molecule templates

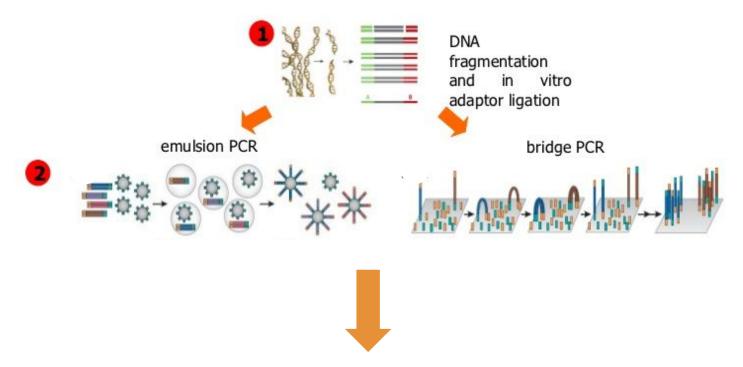


Thousands of primed, single-molecule templates

Ejemplo del proceso general NGS

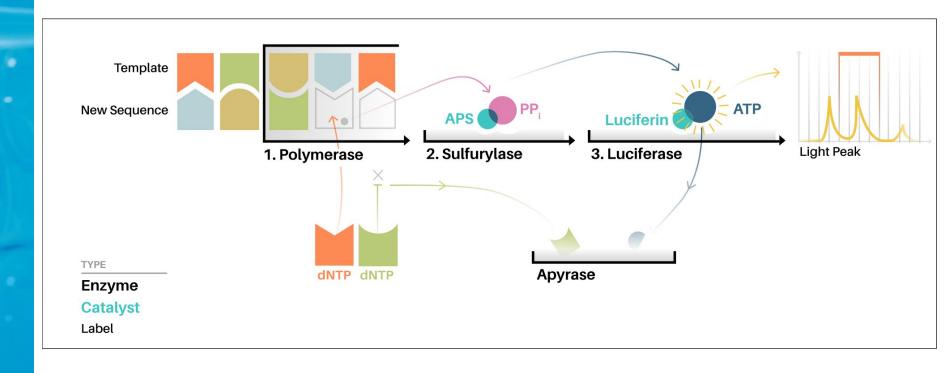


Ejemplo del proceso general NGS

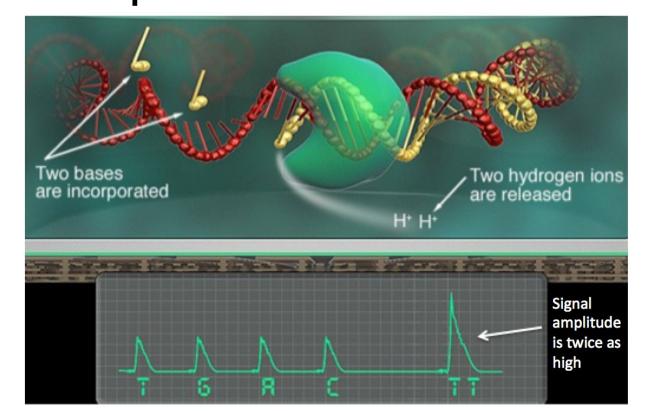


Diferentes métodos de secuenciación

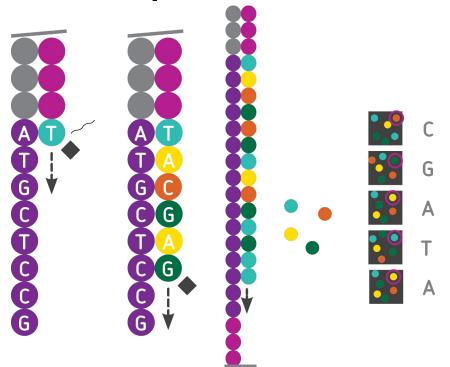
Que pueden ir de lo "complejo"... Pirosecuenciación - 454/Roche



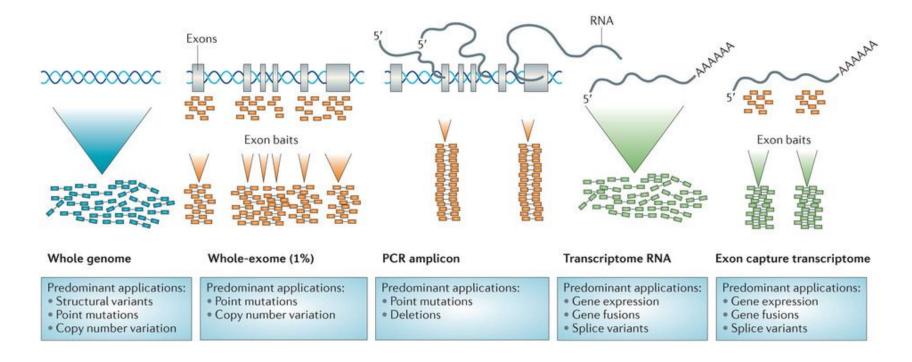
Hasta las más elegantes por su simplicidad... Secuenciación por semiconductores - lon Torrent



Durante el resto del curso nos enfocaremos en la tecnología de Illumina (Secuenciación por terminadores reversibles)



¿Qué es lo que nos interesa secuenciar?



¿Qué es lo que nos interesa secuenciar?

