EXPLORANDO LA TECNOLOGÍA ISCAN DE ILLUMINA A TRAVÉS DEL CHIP GLOBAL SCREENING ARRAY (GSA V2.0)

WHOLE-GENOME GENOTYPING ARRAY

Pelayo González de Lena Rodríguez

1. INTRODUCCIÓN

Históricamente, ha habido dos grandes compañías de genotipificación: Illumina y Affymetrix. Los diseños de los arreglos de estas dos compañías se basan en diferentes químicas (LaFramboise, 2009) y tienen diferentes coberturas (Ha et al. , 2014 ; Jiang et al ., 2013). Affymetrix diseña todas sus matrices de genotipado utilizando la línea directa de referencia. Illumina, por otro lado, adoptó una definición de hebra personalizada no convencional, que puede causar confusión y un análisis posterior innecesariamente complicado (Nelson et al. , 2012 , 2014).

La cadena de un SNP puede verificarse comparando la frecuencia de alelos en el conjunto de datos con informes previos de una base de datos de población apropiada. Sin embargo, cuando la frecuencia alélica está cerca del 50%, todavía puede surgir ambigüedad. Además, el problema de la cadena se puede resolver comparando los alelos con un genoma de referencia. Sin embargo, cuando dos alelos de los SNP son complementarios (A / T o C / G), la verdadera cadena permanece indeterminada. La única solución absoluta para determinar la cadena es comparar las secuencias de la sonda con un genoma de referencia, siempre que las secuencias de la sonda sean correctas. Para la mayoría de los usuarios de la genotipificación de micromatrices, hasta ahora no ha sido práctico acceder a las secuencias de la sonda y compararlas con la secuencia de referencia.

Este análisis tiene como objetivo principal la creación, búsqueda, implementación o uso permitido de aquellas herramientas computacionales que permitan resolver algunos de los problemas más comunes que nos encontramos al producir datos de genotipado de alto rendimiento. En este caso, hemos centrado nuestra atención en el chip Global Screening Array GSA-24v2-0_A1 y en analizar la calidad del manifest propuesto por Illumina, resolver el problema de la orientación de los SNP's para cada hebra.

Por último hemos realizado una prueba sobre dos array de genotipado, uno a partir de datos crudos de intensidades (.idat) y el otro a partir de ficheros de calling (.gtc)

2. INFORMACIÓN SOBRE EL CHIP:

El Global Screening Array Consortium reunió a investigadores en enfermedades humanas e investigación traslacional para diseñar una matriz de genotipado de próxima generación para la genética a escala poblacional, el cribado de variantes y la investigación de medicina de precisión. El contenido de la matriz incluye contenido de genoma multiétnico altamente optimizado, variantes curadas de investigación clínica y marcadores de control de calidad (QC) diseñados para ser útiles en una amplia gama de aplicaciones, poblaciones y enfermedades. El contenido multiétnico del genoma ofrece una alta cobertura genómica y un rendimiento de imputación en 26 poblaciones continentales. El contenido de investigación clínica curada incluye más de 50,000 variantes que incluyen marcadores para farmacogenómica, investigación de cribado neonatal y perfil de riesgo.

Además, hay más de 9,000 QC y marcadores de alto valor que permiten la identificación de la muestra, el seguimiento, la determinación de la ascendencia y la estratificación de gran valor para aplicaciones genéticas y de detección a gran escala. Los datos acumulados resultantes de los miembros del consorcio pueden facilitar una atención de salud más personalizada en el futuro, ya que la genómica se convierte potencialmente en parte de la práctica clínica.

3. INFORMACIÓN SOBRE EL SISTEMA TOP/BOT

Un aspecto del diseño de la matriz de Illumina que siempre ha molestado a los investigadores es el problema de la consistencia de las hebras. Los cromosomas autosómicos son diploides y bicatenarios. En GWAS, los SNP

significativos siempre se informan como alelo de riesgo (AR), ocasionalmente denominado alelo efectivo (EA). El AR se puede presentar en la cadena directa o en la inversa del genoma. Cuando se diseñan las sondas para detectar los dos alelos de un SNP, las sondas pueden diseñarse para la cadena directa o inversa. Para la mayoría de las veces, el AR notificado para un SNP se convertirá en el capítulo directo antes de la publicación. Sin embargo, no podemos confiar completamente en los autores sobre el tema de la coherencia de la cadena. La incoherencia del capítulo dará lugar a una interpretación errónea de la dirección de la asociación,

En el caso de una matriz de genotipado, una simple definición de cadena directa e inversa para cada SNP sería suficiente. En su lugar, Illumina introdujo una definición más clara de hebra, arriba y abajo, que ha causado una gran confusión con la hebra hacia adelante y hacia atrás. Más preocupante, al generar los datos de genotipado de GenomeStudio, se puede seleccionar una opción para convertir todos los SNP a la cadena de reenvío.

Esta opción debería resolver el problema de convertir a las anotaciones estándar hacia adelante y hacia atrás. Sin embargo, calculamos que entre el 1% y el 11% de los SNP no se presentan en la cadena directa después de la conversión de GenomeStudio en varias matrices de genotipificación de Illumina. La cadena de un SNP puede verificarse comparando la frecuencia de alelos en el conjunto de datos con los reportados previamente en una base de datos de población apropiada.

Sin embargo, Cuando la frecuencia del alelo está cerca del 50%, todavía puede surgir ambigüedad. Además, el problema de la cadena se puede resolver comparando los alelos con un genoma de referencia. Sin embargo, cuando los dos alelos de los SNP son complementarios (A / T o C / G), nos quedamos sin la capacidad de determinar la cadena verdadera. La única solución absoluta para este problema es determinar la cadena comparando las secuencias de la sonda con el genoma de referencia, siempre que las secuencias de la sonda sean correctas. Para la mayoría de los usuarios de la genotipificación de micromatrices, hasta ahora no ha sido práctico acceder a las secuencias de la sonda y compararlas con la secuencia de referencia.

Todavía nos quedan sin la capacidad de determinar la verdadera cadena. La única solución absoluta para este problema es determinar la cadena comparando las secuencias de la sonda con el genoma de referencia, siempre que las secuencias de la sonda sean correctas. Para la mayoría de los usuarios de la genotipificación de micromatrices, hasta ahora no ha sido práctico acceder a las secuencias de la sonda y compararlas con la secuencia de referencia. Todavía nos quedan sin la capacidad de determinar la verdadera cadena. La única solución absoluta para este problema es determinar la cadena comparando las secuencias de la sonda con el genoma de referencia, siempre que las secuencias de la sonda sean correctas.

4. DESIGNACIÓN DE HEBRA:

SNPs de Illumina

Los resultados de genotipado de los ensayos Illumina GoldenGate e Infinium se envían en formato A/B de Illumina, y se almacenan como A-allele = A y B-allele = B.

Las llamadas de nucleótidos reales correspondientes a A y B se toman de los archivos de manifiestos de Illumina (ejemplo), interpretados de acuerdo con la documentación de la Nota Técnica de Illumina "TOP/BOT" Strand y "A/B" Allele.

En resumen:

1. Para los SNPs[A/C] y[A/G], Allele A es A.
2. Para los SNPs[T/C] y[T/G], el alelo A es T y la secuencia se denomina BOT.

3. Para SNPs[A/T], cuando la hebra de la Ilumina es TOP entonces alelo A = A y alelo B = T. Cuando la hebra es BOT, entonces alelo A = T y alelo B = A.
4. Para[C/G] SNPs, cuando la hebra de la Ilumina es TOP entonces alelo A = C y alelo B = G. Cuando la hebra es BOT entonces alelo A = G y alelo B = C.

5. PROCESO AUTOMATIZADO DE GENERACIÓN DE FINAL REPORT CON GENOMESTUDIO

Aunque el uso del chip de genotipado GSA v2.0 ofrece nuevas y emocionantes oportunidades, también presenta desafíos adicionales en el procesamiento de datos. GenomeStudio, que se utiliza para agrupar todos los arreglos de genotipado de Illumina, está diseñado para identificar SNP comunes en lugar de SNP raros.

El algoritmo que GenomeStudio usa para agrupar, GenCall, se diseñó originalmente como parte de BeadStudio para los arrays Illumina 550K muchos años antes de la introducción del chip del exoma en 2011. Múltiples estudios han demostrado que el algoritmo GenCall es más adecuado para la identificación de SNP comunes que para los SNP raros.

El uso de GenomeStudio para agrupar los datos de genotipado del chip GSA v2.0 dará como resultado muchos SNP mal agrupados. Existen dos tipos de tales agrupaciones incorrectas: (i) genotipos que se agruparon incorrectamente y, por lo tanto, se les asignó un genotipo incorrecto; y (ii) los genotipos que se configuraron como "faltantes" debido a que el algoritmo GenCall no pudo interpretar una llamada de genotipo. Se requieren nuevas estrategias para corregir estos SNP raros mal agrupados.

Asignar la orientación de hebra de un SNP siempre ha sido un reto para los investigadores en genética. Las diferencias en la orientación de los filamentos pueden causar confusión al comparar los resultados en múltiples estudios y plataformas. Comúnmente se implementan cuatro convenciones de nomenclatura de cadenas: sonda-objetivo, más-menos, adelante-atrás y arriba-abajo.

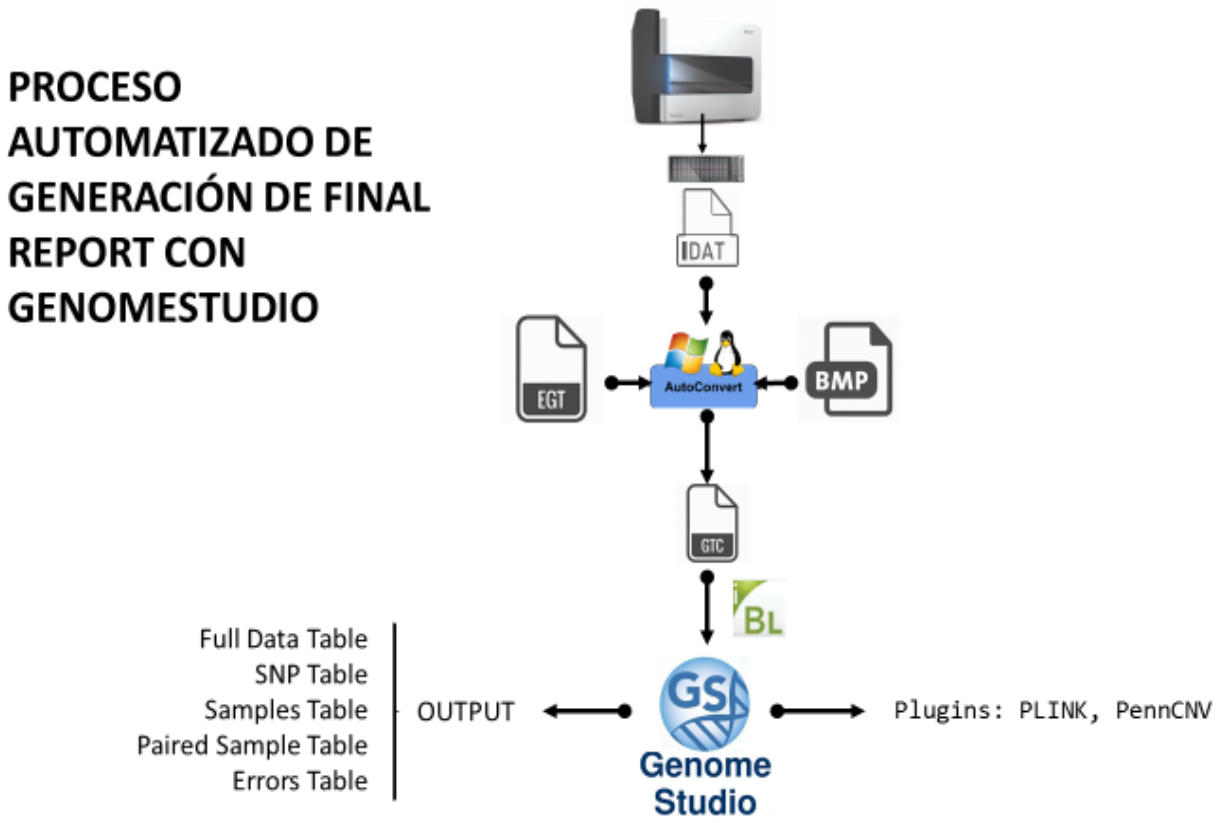
La convención de arriba a abajo fue desarrollada por Illumina, y se usa en todas las matrices de genotipado de Illumina. La idea detrás del sistema de orientación de cadena superior-inferior es permitir que los sistemas de Illumina designen constantemente la misma orientación de SNP y llamada alélica, incluso si la base de datos de SNP (dbSNP) o la referencia humana cambian. Si bien esta idea tiene sus méritos, no se la comprende ampliamente. Las reglas para determinar las cadenas superior e inferior son las siguientes: un SNP está en la cadena superior si el primer alelo es A y el segundo alelo es C o G; un SNP está en la cadena inferior si el primer alelo es T y el segundo alelo es C o G. Sin embargo, en cuanto a los SNP A / T y C / G la definición de cadenas sigue siendo ambigua, Illumina introdujo una técnica de "secuencia de pasos" para designar la orientación de hebra y alelo para los SNP A / T y C / G 15.

La técnica de caminar en secuencia funciona de la siguiente manera: deje que la posición del SNP ambiguo sea n , y las posiciones antes y después de n se pueden denotar como n - 1, n - 2, ... y n + 1, n + 2 ..., respectivamente. Dos caminantes se alejarán de n hacia la izquierda y hacia la derecha, un nucleótido a la vez. Así, se observa un nuevo par de nucleótidos cada vez (el primero sería (n - 1 / n+ 1)). Si se observa una A o T en el primer par no ambiguo en el lado 5 'del SNP, entonces se define que el SNP está en la cadena superior; si se observa A o T en el primer par no ambiguo en el lado 3 'del SNP, entonces se define que el SNP está en la cadena inferior. Sin embargo, incluso cuando Illumina puso en práctica este nuevo esquema de hilos, se siguieron generando numerosos errores de filamentos.

Al exportar datos de SNP desde GenomeStudio, una opción permite cambiar todas las orientaciones de cadena de SNP a 'reenviar', definidas por la cadena de avance registrada en el dbSNP. Existen algunas diferencias menores entre las definiciones de las cadenas 'dbSNP' forward 'y HG19' plus '.

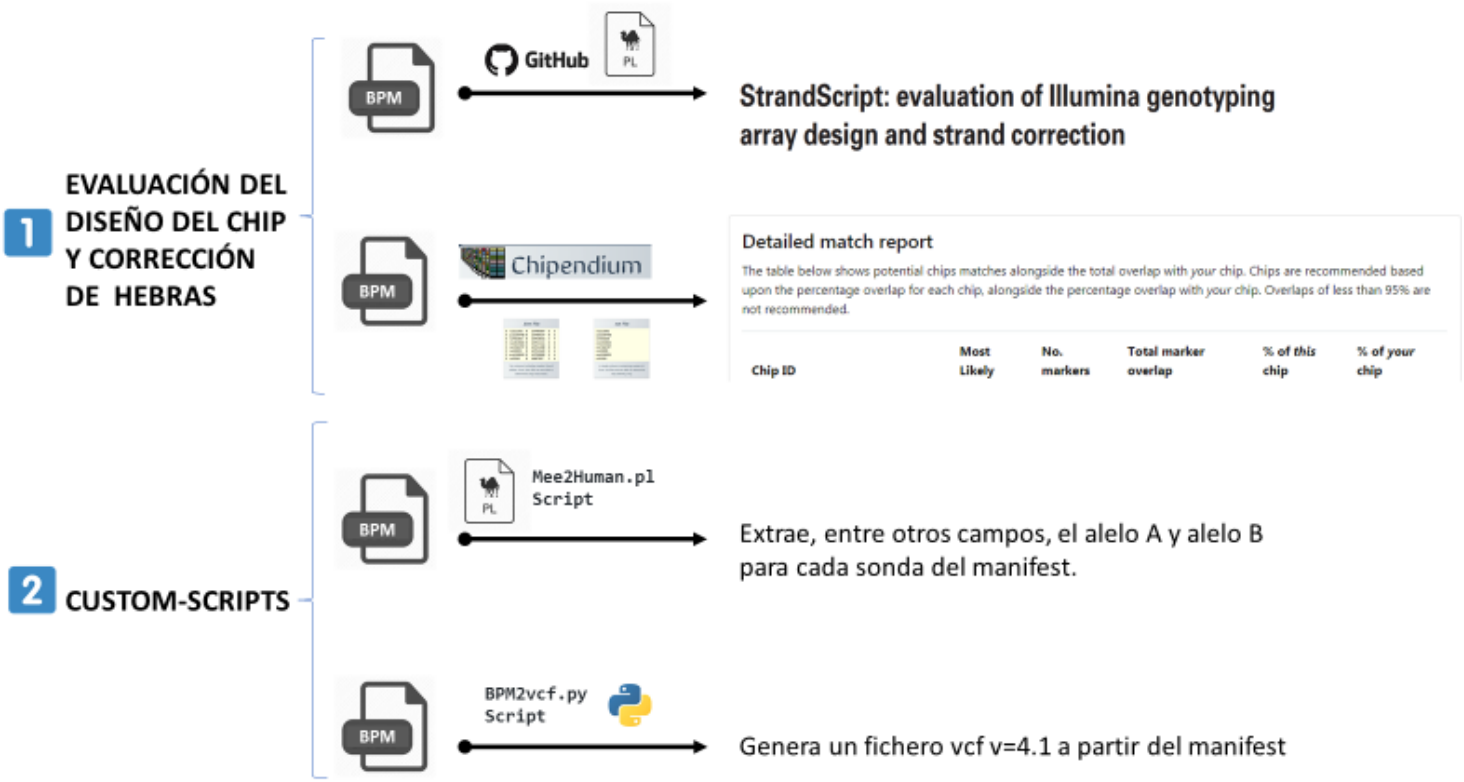
Después de convertir a dbSNP-forward, 2,055 SNPs no están en la misma cadena que HG19-plus. La mayoría de estos (excepto 35 SNPs; Cuadro 2 complementario) se puede convertir a la cadena HG19-plus utilizando el método descrito en https://www.well.ox.ac.uk/~wrayner/strand/

GenomeStudio muestra agrupamientos en dos formatos: coordenadas polares y coordenadas cartesianas. Las coordenadas cartesianas muestran el grupo utilizando los valores de intensidad normalizados A y B (que denotan los dos alelos específicos). Las coordenadas polares muestran el grupo utilizando los valores R normalizados y θ normalizados para denotar los ejes y y x , respectivamente.



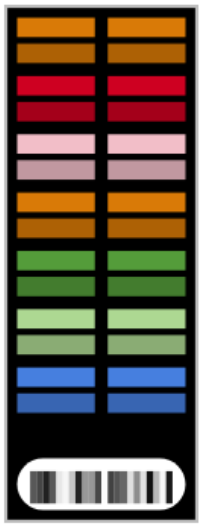
Minimum GenomeStudio System Recommendations	
CPU Speed	2.0 GHz or greater
Processor	2 or more cores
Memory	8 GB or more
Hard Drive	250 GB or larger
Operating System	Windows 7 or higher

6. CONTROL DE CALIDAD DEL MANIFEST FILE



1. EVALUACIÓN DEL DISEÑO DEL CHIP Y CORRECCIÓN DE HEBRAS

   a. Chipendium: Chipendium puede identificar la plataforma de microarrays más probable y la cadena actual de sus metadatos de chip Illumina. Conocer la plataforma y la cadena correctas es esencial para usar el sitio web de Will Rayner llamado Strand que usamos a continuación. Chipendium necesita un fichero .BIM o un .TXT (col2toChipendium.txt) con todos los ID SNP's disponibles en el chip.



Your chip data

Filename	col2toChipedium.txt
File type	TXT
No. markers	665.633

Detailed match report

The table below shows potential chips matches alongside the total overlap with *your* chip. Chips are recommended based upon the percentage overlap for each chip, alongside the percentage overlap with *your* chip. Overlaps of less than 95% are not recommended.

Chip ID	Most Likely	No. markers	Total marker overlap	% of *this* chip	% of *your* chip
GSA-24v2-0_A2	✓	665.608	665.608	100.00	100.00
GSA-24v2-0_A1	✓	665.608	665.608	100.00	100.00
GSCA-24v2-0_20023496_A1	✗	713.372	648.363	90.89	97.41
GSAMD-24v2-0_20024620_A1	✗	759.993	648.364	85.31	97.41

b. StandScript: https://github.com/seasky002002/Strandscript.

Dado que es posible que un SNP se diseñe correctamente en una matriz, pero que el archivo de manifiesto contenga errores, o viceversa, StrandScript tiene la capacidad de examinar el archivo de manifiesto de Illumina para identificar SNP con diseños potencialmente erróneos.

Aunque la tasa de error dentro del archivo de manifiesto para cada una de las principales matrices de genotipificación de Illumina es solo una fracción del diseño completo (0.003–0.22%) basado en nuestro análisis, el filtrado de estos SNP potencialmente problemáticos aumentará la integridad general de los datos.

Paso 1: Comprobación del archivo de manifiesto de Illumina (entrada: manifest (.bpm))

El paso 1 envía dos archivos al directorio de salida. Genera dos archivos (new_manifest.csv & outdated_manifest.csv) en la carpeta test/.

new_manifest.csv: lista la información básica y el número de pares de bases mal asignados para snps. Este archivo se proporcionaría como entrada para el paso 2.

obsoleto_manifest.csv: lista los snps obsoletos.

En nuestro caso:

1232 SNP's no están actualmente actualizados o bien mapeados contra la v del hg19 que usamos (b3).

```
==> outdated_GSA-24v2-0_A1.csv <==
exm2262791-0_B_R_1975244884,exm2262791,[T/G],37,XY,2594011
exm2273223-0_T_R_1984849553,exm2273223,[A/G],37,XY,181779
exm2273224-0_B_F_1984849370,exm2273224,[T/C],37,XY,1429155
ilmnseq_rs1038470-138_T_F_2602852762,rs1038470,[A/T],37,XY,90956204
ilmnseq_rs11096456-147_T_F_2602852738,rs11096456,[C/G],37,XY,89225502
ilmnseq_rs112096861-147_B_R_2602852721,rs112096861,[T/C],37,XY,91796265
ilmnseq_rs113306384-147_B_R_2602852711,rs113306384,[T/G],37,XY,91782556
ilmnseq_rs12557859-147_T_F_2602852682,rs12557859,[A/G],37,XY,92102646
ilmnseq_rs1369727-138_T_F_2602852672,rs1369727,[A/G],37,XY,90247280
ilmnseq_rs1435909-147_B_F_2602852670,rs1435909,[T/C],37,XY,90274598
```

```
(base) pelayo@biopelayo-SATELLITE-C50D-B:~/Escritorio/ICM/Illumina_data_GSA_V2.0$
wc -l outdated_GSA-24v2-0_A1.csv
1233 outdated_GSA-24v2-0_A1.csv
(base) pelayo@biopelayo-SATELLITE-C50D-B:~/Escritorio/ICM/Illumina_data_GSA_V2.0$
wc -l new_GSA-24v2-0_A1.csv
664376 new_GSA-24v2-0_A1.csv
```

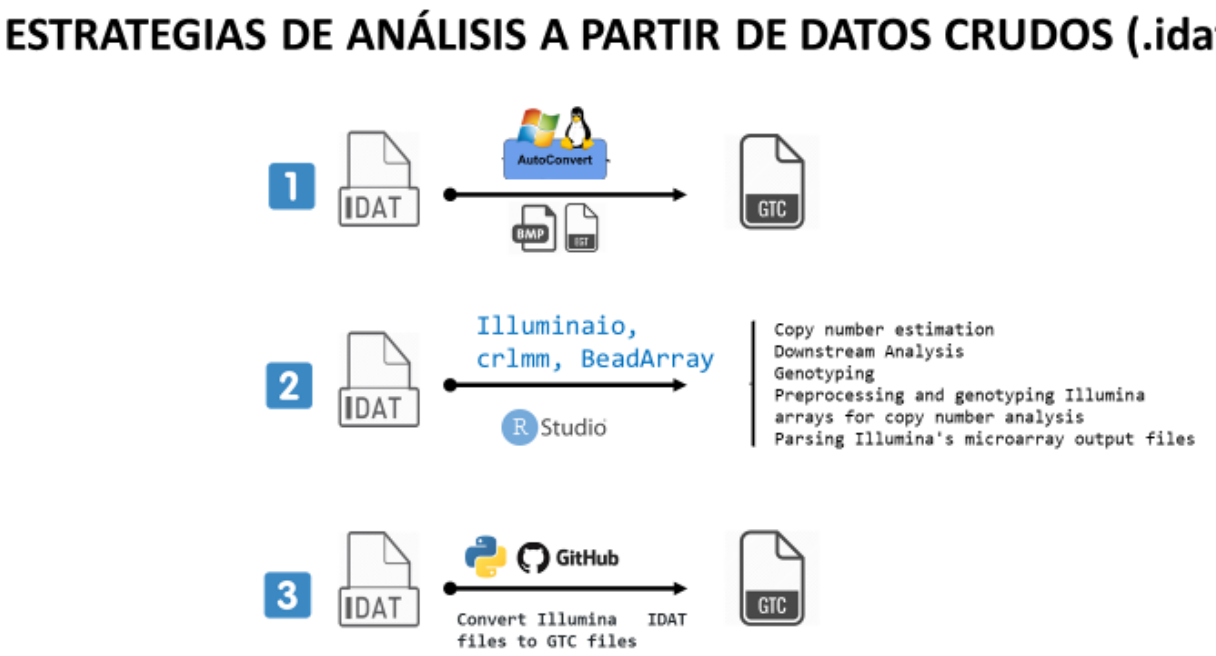
2. CUSTOM-SCRIPTS

Hemos generado dos scripts: uno el perl y otro en python que sean capaces de reconocer los alelos a y b del manifest y otro que genere un fichero vcf 4.1 a partir del manifest file.

1. Mee2Human.pl: Extrae alelos
2. BPM2vcf.py : Genera .vcf a partir de .bpm

7. ESTRATEGIAS DE ANÁLISIS A PARTIR DE DATOS CRUDOS (.idat)

1. Vía AutoConvert y .Mono. Nos descargamos el software y la librería para no tener que utilizar un servidor Windows y hacemos el calling igual, obteniendo los mismos resultados de calidad para todos los ficheros.



2. Vía Rtsudio. Existen varias librerías disponibles actualizadas para usar con ficheros .idat de genotipado

3. Vía Python: Utilizamos la herramienta gtc2vcf.py que posee una función específica para parsear ficheros de intensidad .idat y generar el calling en formato .gtc

- Generación de archivos GTC

Si tiene archivos de datos de intensidad (IDATs) para los cuales los archivos GTC no están disponibles, es posible generar manualmente estos archivos con el software Beeline o AutoConvert (https://support.illumina.com/array/array_software/beeline/downloads.html).

Beeline proporciona una interfaz gráfica de usuario para la creación de archivos GTC, mientras que AutoConvert permite generar archivos GTC desde la línea de comandos.

8. ESTRATEGIAS DE ANÁLISIS A PARTIR DE GenTrain/GenCall (.gtc)

1. IlluminaBeadArrayFiles

Mediante el uso de dos funciones (gtc_final_report.py + locus_summary.py) en python generamos un fichero output y otro llamado outputLocus para cada uno de los arrays. Presentan este aspecto:

```
pelayo@biopelayo-SATELLITE-C50D-B:~/Escritorio/ICM/scripts/BeadArrayFiles-
develop/examples$ more output
[Header]
Processing Date 02/14/2019 04:22 PM
Content GSA-24v2-0_A1.bpm
Num SNPs        665608
Total SNPs      665608
Num Samples     17
Total Samples   17
[Data]
SNP Name        Sample ID       Chr     MapInfo Alleles - AB    Alleles - Plus
        Alleles - Forward
1:103380393     202502040079_R09C02     1       103380393       BB      GG      GG
1:109439680     202502040079_R09C02     1       109439680       AA      AA      AA
1:118227370     202502040079_R09C02     1       118227370       AA      TT      TT
```

```
pelayo@biopelayo-SATELLITE-C50D-B:~/Escritorio/ICM/scripts/BeadArrayFiles-
develop/examples$ more outputLocus120
Locus Summary on /home/pelayo/Escritorio/ICM/scripts/BeadArrayFiles-
develop/examples/outputLocus120
LOCI = 665608,#
DNAs = 24,
ProjectName = Project,GenCall
Version = < 6.1.3.0,Low
GenCall Score Cutoff = NaN

Row,Locus_Name,Illumicode_Name,#No_Calls,#Calls,Call_Freq,A/A_Freq,A/B_Freq,B/B_Fr
eq,Minor_Freq,Gentrain_Score,50%_GC_Score,10%_GC_Score,Het_Excess_Fr
eq,ChiTest_P100,Cluster_Sep,AA_T_Mean,AA_T_Std,AB_T_Mean,AB_T_Std,BB_T_Mean,BB_T_S
td,AA_R_Mean,AA_R_Std,AB_R_Mean,AB_R_Std,BB_R_Mean,BB_R_Std,Plus/Min
us Strand
```

BeadArrayFiles es un librería para analizar los formatos de archivo relacionados con las matrices de perlas de Illumina. Los archivos GTC son producidos por los softwares AutoConvert y AutoCall y contienen información de genotipado (y de otro tipo) codificada en un formato binario.

La biblioteca IlluminaBeadArrayFiles proporciona un analizador para extraer información de estos archivos binarios.

2. GTCtoVCF: Herramienta propuesta desde el GitHub de Illumina y desarrollada por ellos en python.

GTC to VCF converter

Usage

usage: gtc_to_vcf.py [-h] [--gtc-paths GTC_PATHS [GTC_PATHS ...]]

--manifest-file MANIFEST_FILE --genome-fasta-file

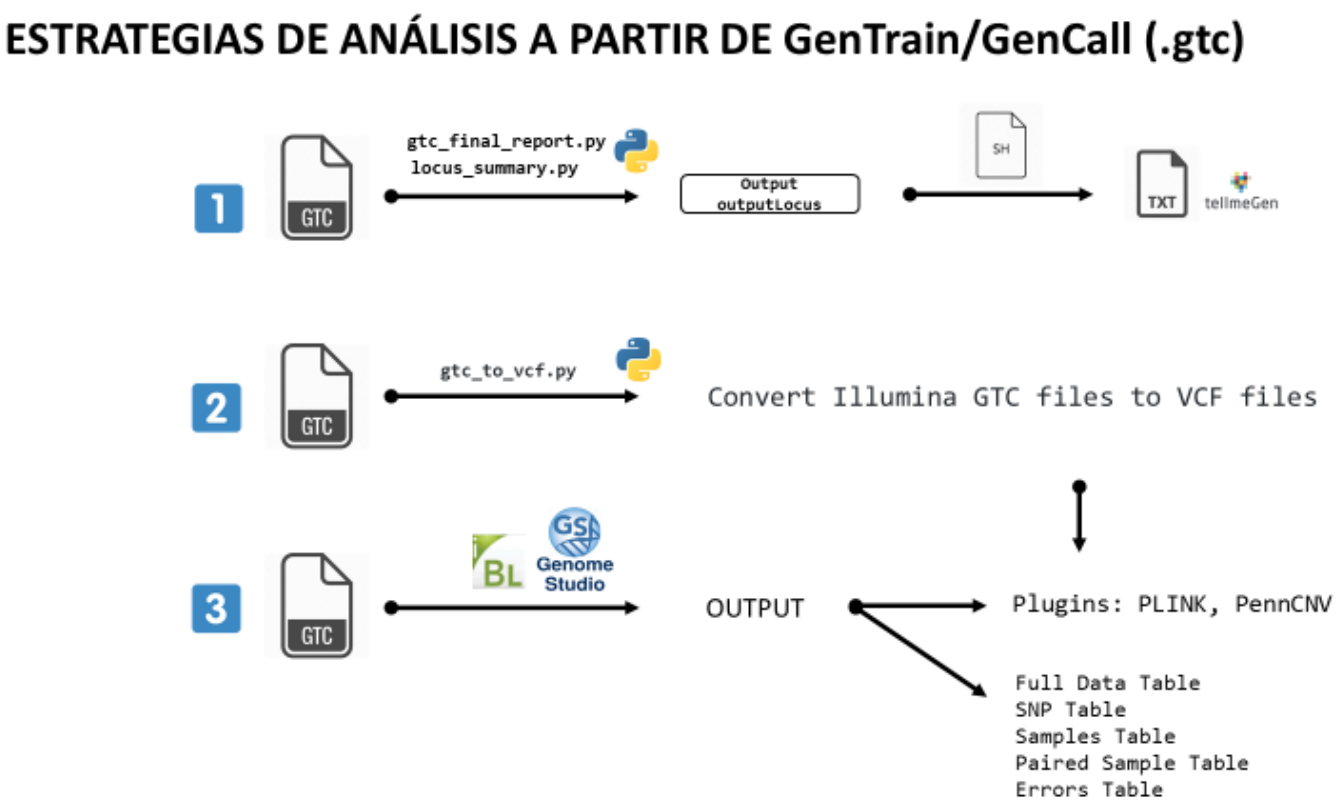
GENOME_FASTA_FILE [--output-vcf-path OUTPUT_VCF_PATH]

[--skip-indels] [--log-file LOG_FILE]

[--expand-identifiers] [--unsquash-duplicates]

[--auxiliary-loci AUXILIARY_LOCI]

[--filter-loci FILTER_LOCI] [--disable-genome-cache]



3. Obtención de ficheros .vcf a partir de ficheros .gtc

Illumina proporciona el software Beeline de forma gratuita y esto incluye el ejecutable AutoConvert que permite llamar genotipos a partir de datos de intensidad cruda utilizando el algoritmo GenCall propietario de Illumina. AutoConvert está casi totalmente escrito en lenguaje Mono/.Net, con la excepción de una pequeña función matemática (findClosestSitesToPointsAlongAxis) que está contenida en una biblioteca de Windows PE32+ (MathRoutines.dll).

Como se trata de un código no gestionado, para ejecutarse en Linux con Mono necesita estar incrustado en una biblioteca equivalente de Linux ELF64 (libMathRoutines.dll.so) como se muestra a continuación. Esta función se ejecuta como parte de la normalización de las intensidades brutas cuando se toman muestras de 400 homocigotos candidatos antes de llamar a los genotipos.

Por algunas razones poco claras, también necesitará descargar por separado una biblioteca Mono/.Net adicional (Heatmap.dll) de GenomeStudio e incluirla en su directorio binario como se muestra a continuación, muy probablemente debido a las diferencias en las que Mono y .Net resuelven las dependencias de las bibliotecas.

Las especificaciones para los archivos BPM, EGT y GTC de Illumina se obtuvieron a través de la biblioteca BeadArrayFiles de Illumina y el script GTCtoVCF. Las especificaciones para los archivos IDAT se obtuvieron a través del paquete illuminaio de Henrik Bengtsson. La determinación de la cadena de referencia se realiza utilizando la asignación de cadenas TOP/BOT de Illumina en el archivo de manifiesto. El plugin gtc2vcf bcftools es cientos de veces más rápido que el script de Illumina y puede ser usado para convertir archivos GTC a VCF.

Descripción de la salida

La salida del archivo VCF sigue el formato VCF4.1 (https://samtools.github.io/hts-specs/VCFv4.1.pdf). Algunos detalles adicionales sobre el formato de salida:

Los genotipos se ajustan para reflejar la ploidía de la muestra. Las llamadas son haploides para los loci en los cromosomas Y, MT y X no PAR para los hombres.

Múltiples SNPs en el manifiesto de entrada que están mapeados a la misma coordenada cromosómica (por ejemplo, loci tri paralelos o sitios duplicados) se colapsan en una entrada VCF y se genera un genotipo combinado.

Para producir el genotipo combinado, se enumera el conjunto de todos los genotipos posibles basándose en los alelos interrogados.

Se filtran los genotipos que no son posibles en base a los llamados alelos y las limitaciones de diseño del ensayo (por ejemplo, los diseños infiniumII no pueden distinguir entre llamadas A/T y C/G). De lo contrario, el genotipo es ambiguo (más de 1) o inconsistente (menos de 1) y se devuelve una llamada no solicitada.

* Convertir el informe final de Illumina GenomeStudio a VCF

Alternativamente, si en su lugar se proporciona un informe final de GenomeStudio en formato matricial, y sigue la siguiente convención (Illumina no comparte especificaciones para este formato de archivo):

```
Chromosome      Position IlmnStrand      SNP     Name    SM.GType SM.Score SM.Theta SM.R
        SM.B Allele Freq SM.Log R Ratio
9       139906359       BOT     [T/C]   200003  AA      0.9299  0.029   1.300   0.0027
        0.2150
9       139926402       TOP     [A/G]   200006  AB      0.7877  0.435   2.675   0.4742
        0.3024
2       220084902       BOT     [T/C]   200047  AA      0.8612  0.083   0.476   0.0532
        -0.1173
2       220089685       TOP     [C/G]   200050  BB      0.8331  0.995   1.499   1.0000
        0.3068
```

9. WORKFLOW DE ANALYSIS PROPUESTO

WORKFLOW DE ANALYSIS PROPUESTO: HETEROGENEIDAD DE FICHEROS

- TODOS LOS FICHEROS DEL PRIMER ARRAY SON .IDAT =24 MUESTRAS/FICHEROS
- TODOS LOS FICHEROS DEL SEGUNDO ARRAY SON .GTC=17 MUESTRAS/FICHEROS

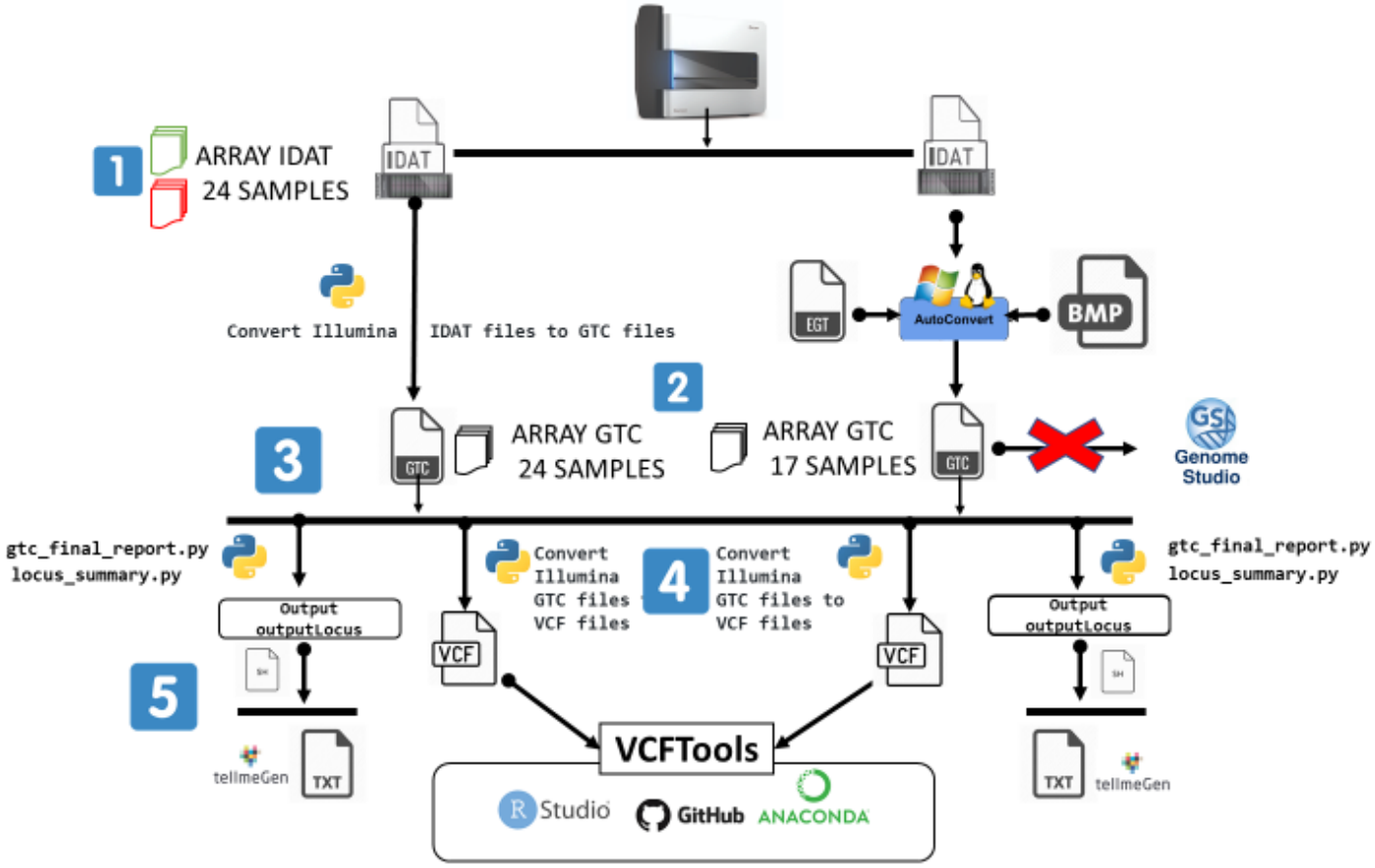
1 OBTENER GTC A PARTIR DE IDAT PARA EL PRIMER ARRAY

2 UNIR TODOS LOS GTC=41 FICHEROS=N

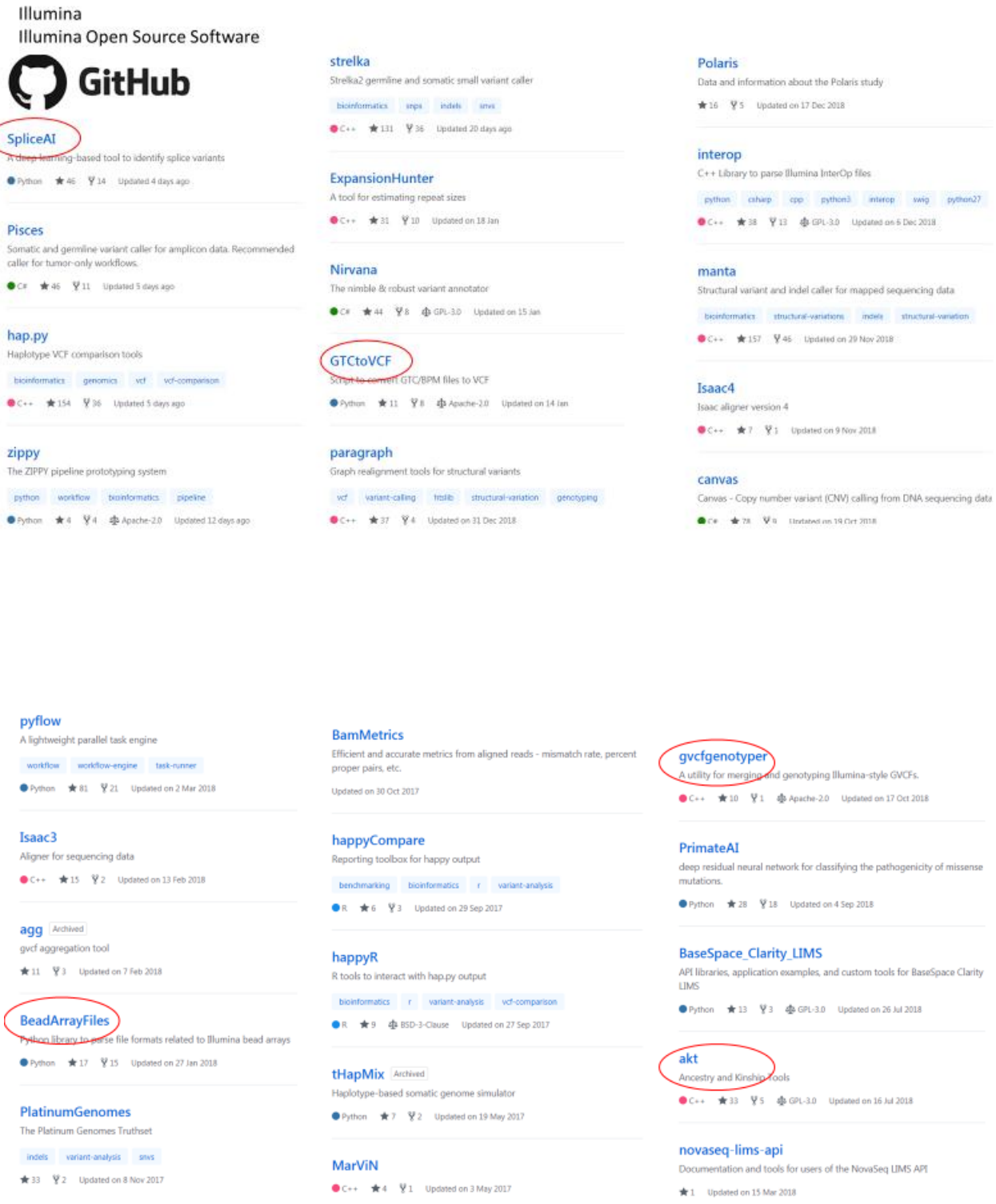
3 CHECK SI EL RESULTADO DE CALIDAD DE LOS GTC ES EL MISMO INDEPENDIENTEMENTE DE LA ESTRATEGIA UTILIZADA.

4 OBTENER UN FICHERO VCF PARA CADA GTC CON INFO SOBRE GENOTIPO + GenCall Score

5 OBTENER GTC REPORT DE TODOS LOS FICHEROS (output) E INFORMACIÓN DETALLADA POR LOCUS PARA TODAS LAS MUESTRAS (N =41)



10. SOFTWARE DE DESARROLLO Y DE LIBRE USO DE ILLUMINA DE DONDE HEMOS OBTENIDO LA MAYORÍA DE INFORMACIÓN PARA ESTE ANÁLISIS




Experimental Biologist
Computational Biologist
NO EATING
NO DRINKING
NO BREATHING
NO OPEN TOE SHOES
NO SHORTS
NO MINI SKIRTS
ALWAYS LAB COATS ON
LIFE IS BEAUTIFUL
www.biocomicals.com