Cuantificar la Metilación del ADN

Análisis de Secuencias
Máster en Genética y Evolución

Guillermo Barturen
gbarturen@ugr.es

Técnicas para cuantificar la metilación del ADN

Pretreatment	Analytical step			
	Locus-specific analysis	Gel-based analysis	Array-based analysis	NGS-based analysis
Enzyme digestion	• *Hpa*II-PCR	• Southern blot • RLGS • MS-AP-PCR • AIMS	• DMH • MCAM • HELP • MethylScope • CHARM • MMASS	• Methyl–seq • MCA–seq • HELP–seq • MSCC
Affinity enrichment	• MeDIP-PCR		• MeDIP • mDIP • mCIP • MIRA	• MeDIP–seq • MIRA–seq
Sodium bisulphite	• MethyLight • EpiTYPER • Pyrosequencing	• Sanger BS • MSP • MS-SNuPE • COBRA	• BiMP • GoldenGate • Infinium	• RRBS • BC–seq • BSPP • WGSBS

AIMS, amplification of inter-methylated sites; BC–seq, bisulphite conversion followed by capture and sequencing; BiMP, bisulphite methylation profiling; BS, bisulphite sequencing; BSPP, bisulphite padlock probes; CHARM, comprehensive high-throughput arrays for relative methylation; COBRA, combined bisulphite restriction analysis; DMH, differential methylation hybridization; HELP, *Hpa*II tiny fragment enrichment by ligation-mediated PCR; MCA, methylated CpG island amplification; MCAM, MCA with microarray hybridization; MeDIP, mDIP and mCIP, methylated DNA immunoprecipitation; MIRA, methylated CpG island recovery assay; MMASS, microarray-based methylation assessment of single samples; MS-AP-PCR, methylation-sensitive arbitrarily primed PCR; MSCC, methylation-sensitive cut counting; MSP, methylation-specific PCR; MS-SNuPE, methylation-sensitive single nucleotide primer extension; NGS, next-generation sequencing; RLGS, restriction landmark genome scanning; RRBS, reduced representation bisulphite sequencing; –seq, followed by sequencing; WGSBS, whole-genome shotgun bisulphite sequencing.

Review

Nature Reviews Genetics **11**, 191-203 | doi:10.1038/nrg2732

ARTICLE SERIES: Applications of next-generation sequencing

Principles and challenges of genome-wide DNA methylation analysis

Peter W. Laird[1] About the author

Técnicas para cuantificar la metilación del ADN

Technology	Features							Potential sources of bias							
	Unambiguous identification of CpG measured	In *cis* co-methylation information	Non-CpG methylation information	Allele-specific measurement capability	Good coverage of regions with low CpG density	Compatible with low amounts of input DNA	Full repeat-masked genome coverage	Copy-number variation bias	Fragment size bias	Incomplete bisulphite conversion bias	Bisulphite PCR bias	Cross-hybridization bias	DNA methylation status bias	GC content bias	CpG density bias
Infinium	(•)					•				•	•	•			
Enzyme–chip	(•)	(•)			(•)				•			•		•	
MeDIP–chip							•	•				•		•	•
BSPP	•	•	•	•						•	•		•		
BC–seq	•	•	•	•						•	•		•		
RRBS	•	•	•	•		•				•	•				
Enzyme–seq	•	•		•	(•)	•			•						
MeDIP–seq				•			•	•						•	•
WGSBS	•	•	•	•	•	•	•			•	•				

'•' indicates that the method has this feature or potentially has this bias; '(•)' indicates that the method has this feature to a limited extent or in some circumstances. BC–seq, bisulphite conversion followed by capture and sequencing; BSPP, bisulphite padlock probes; –chip, followed by microarray; MeDIP, methylated DNA immunoprecipitation; RRBS, reduced representation bisulphite sequencing; –seq, followed by sequencing; WGSBS, whole-genome shotgun bisulphite sequencing.

Review

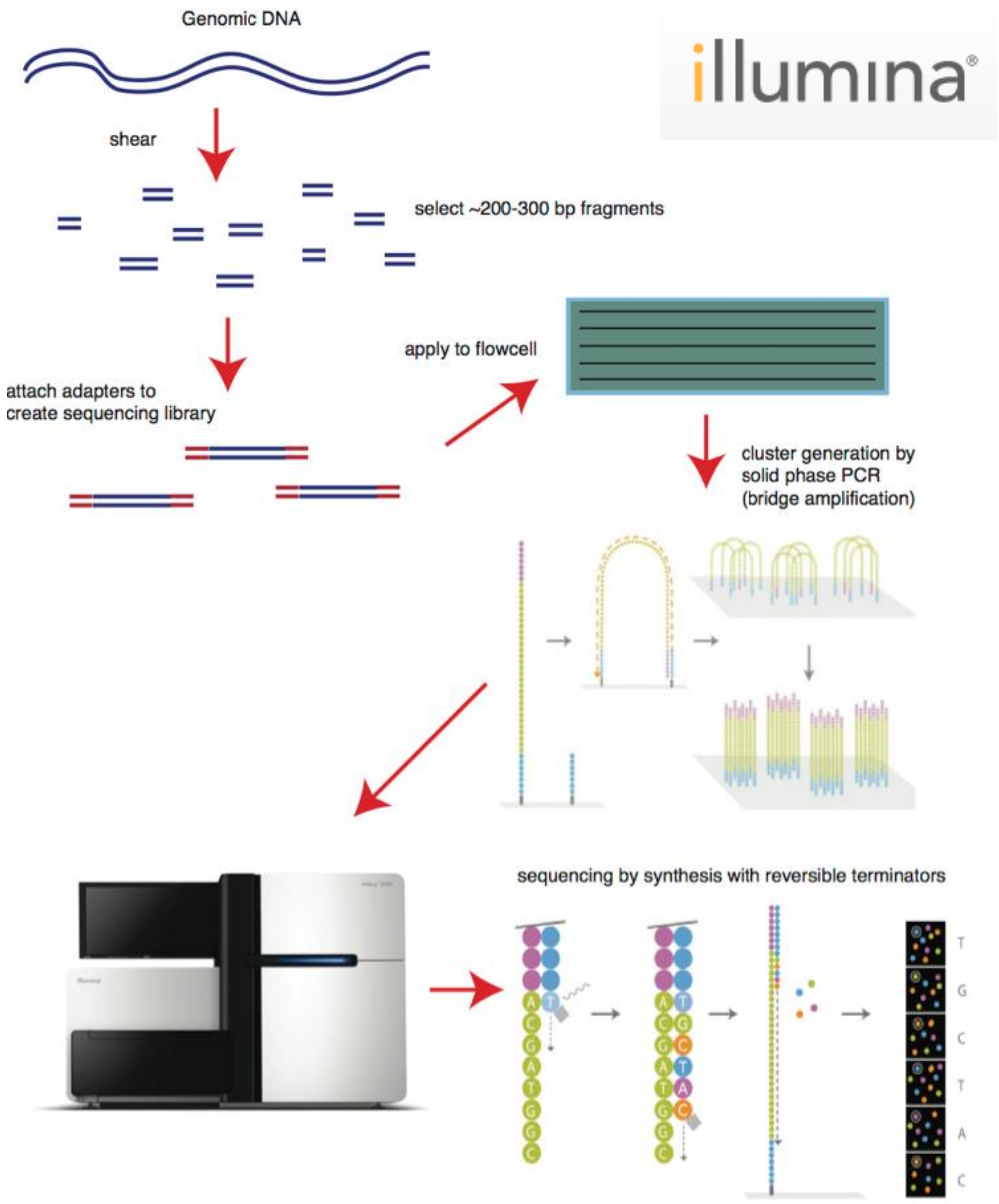
Nature Reviews Genetics **11**, 191-203 | doi:10.1038/nrg2732

ARTICLE SERIES: Applications of next-generation sequencing

Principles and challenges of genome-wide DNA methylation analysis

Peter W. Laird[1] About the author

Secuenciación masiva



Secuenciación masiva

Technology	Sequencing platform	Read length (bp)	Data output	Run time	Recommended application
Illumina	NovaSeq 6000 System		2.4–3.0 Tb	44 h	WGS, WES, PGS
	NextSeq 550 System	150 PE*	100–200 Gb	29 h	
	HISeq 3000/4000 System		up to 1.5 Tb	4 days	
	HiSeq X Series		1.6–1.8 Tb	3 days	
Ion Torrent	Ion GeneStudio S5 System	200 SE	10–15 Gb	19 h	
	Ion GeneStudio S5 Plus System	400 SE	20–30 Gb	10 h	WGS, WES, PGS
	Ion GeneStudio S5 Prime System	200 SE	40–50 Gb	12 h	
	Ion PGM 314 System	400 SE	30–50 Mb	2.3 h	
	Ion PGM 316 System	600 SE	60–100 Mb	3.7 h	
	Ion PGM 318System	200 SE	300–600 Mb	3.0 h	
	Ion Proton System (Ion PI Chip)	200 SE	600 Mb–1 Gb	4.9 h	
		400 SE	600 Mb–1 Gb	4.4 h	
		200 SE	1.2–2 Gb	7.3 h	
		400 SE	up to 15 Gb	2.5 h	
		200 SE			
		400 SE			
		up to 200			
BGI/MGI	DNBSEQ-T7	100 PE, 150 PE	6 Tb	24 h	WGS, WES, PGS
	DNBSEQ-G400/MGISEQ 2000/BGISEQ 500	400 SE, 100 PE, 150 PE,	18.75–1,080 Gb	~78 h	WGS, WES
	DNBSEQ-G400 FAST	200 PE	330 Gb	12–13 h	PGS
	DNBSEQ-G50/MGISEQ 200/BGISEQ 50	100 SE, 150 PE	10–150 Gb	10–64 h	PGS
		50 SE, 100 SE, 50 PE, PE100			

The sequencing performance specifications are as per company description.
bp, base pair; WGS, whole genome sequencing; WES, whole exome sequencing; PGS, targeted generation sequencing; SE, Single-End reads; PE, Paired-End reads; Gb, Gigabytes; Tb, Terabytes.
**Applies to all sequencers.*

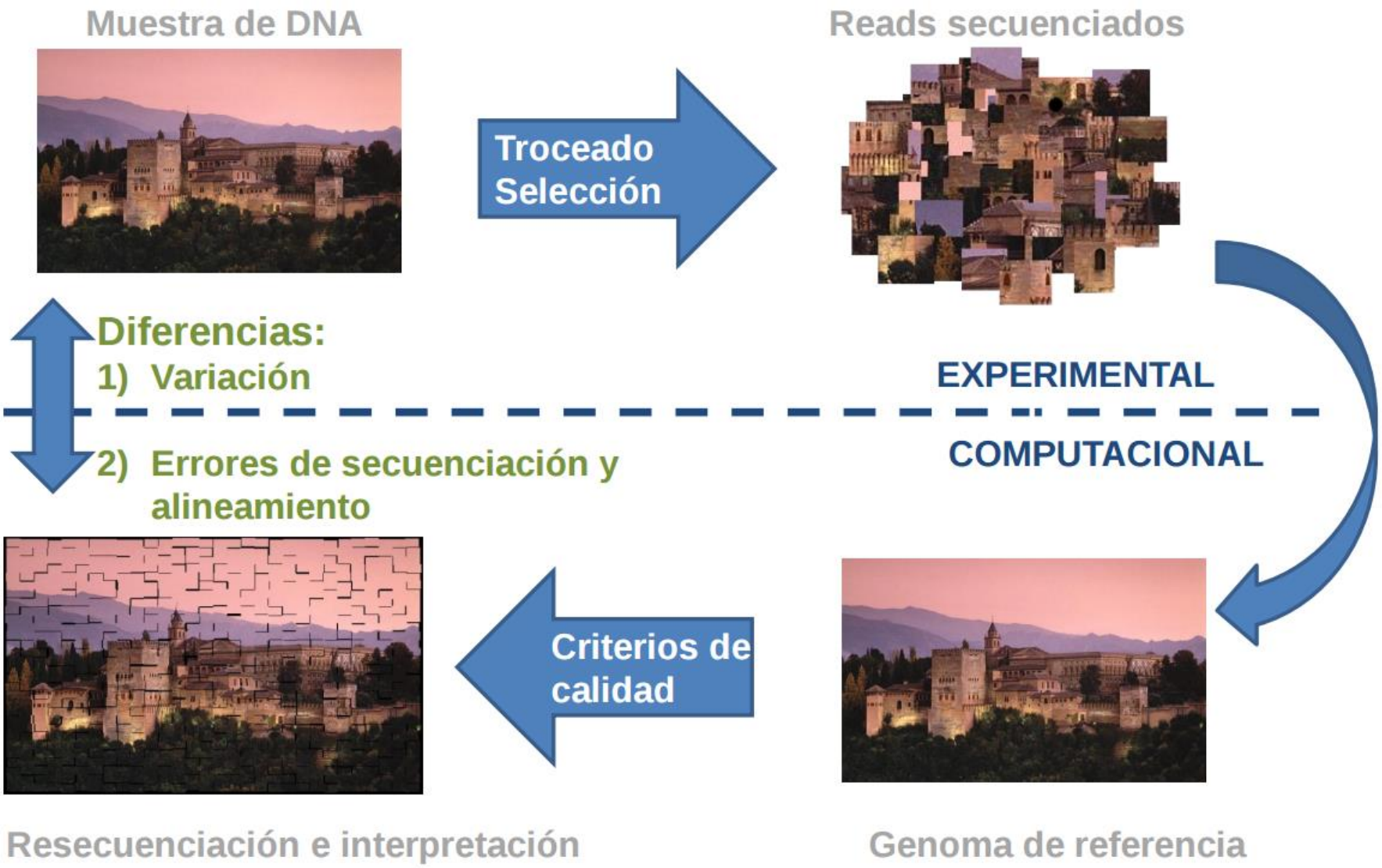
REVIEW article
Front. Genet., 23 October 2020
Sec. Human and Medical Genomics
https://doi.org/10.3389/fgene.2020.544162

Next Generation Sequencing and Bioinformatics Analysis of Family Genetic Inheritance

Aquillah M. Kanzi*, James Emmanuel San, Benjamin Chimukangara, Eduan Wilkinson, Maryam Fish, Veron Ramsuran and Tulio de Oliveira

Kwazulu-Natal Research and Innovation Sequencing Platform (KRISP), School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

Re-Secuenciación masiva



Secuenciación masiva con tratamiento de bisulfito (WGSBS)

Fundamentos

- Tratamiento del ADN con bisulfito sódico
  Citosina > Uracilo > Timina
  m5Citosina > Citosina
- Secuenciación masiva
  El ADN se fragmenta de manera aleatoria y se secuencia
- Comparar fragmentos secuenciados ("*reads*") con referencia
  Desemparejamiento C/T > No Metilación
  Emparejamiento C/C > Metilación

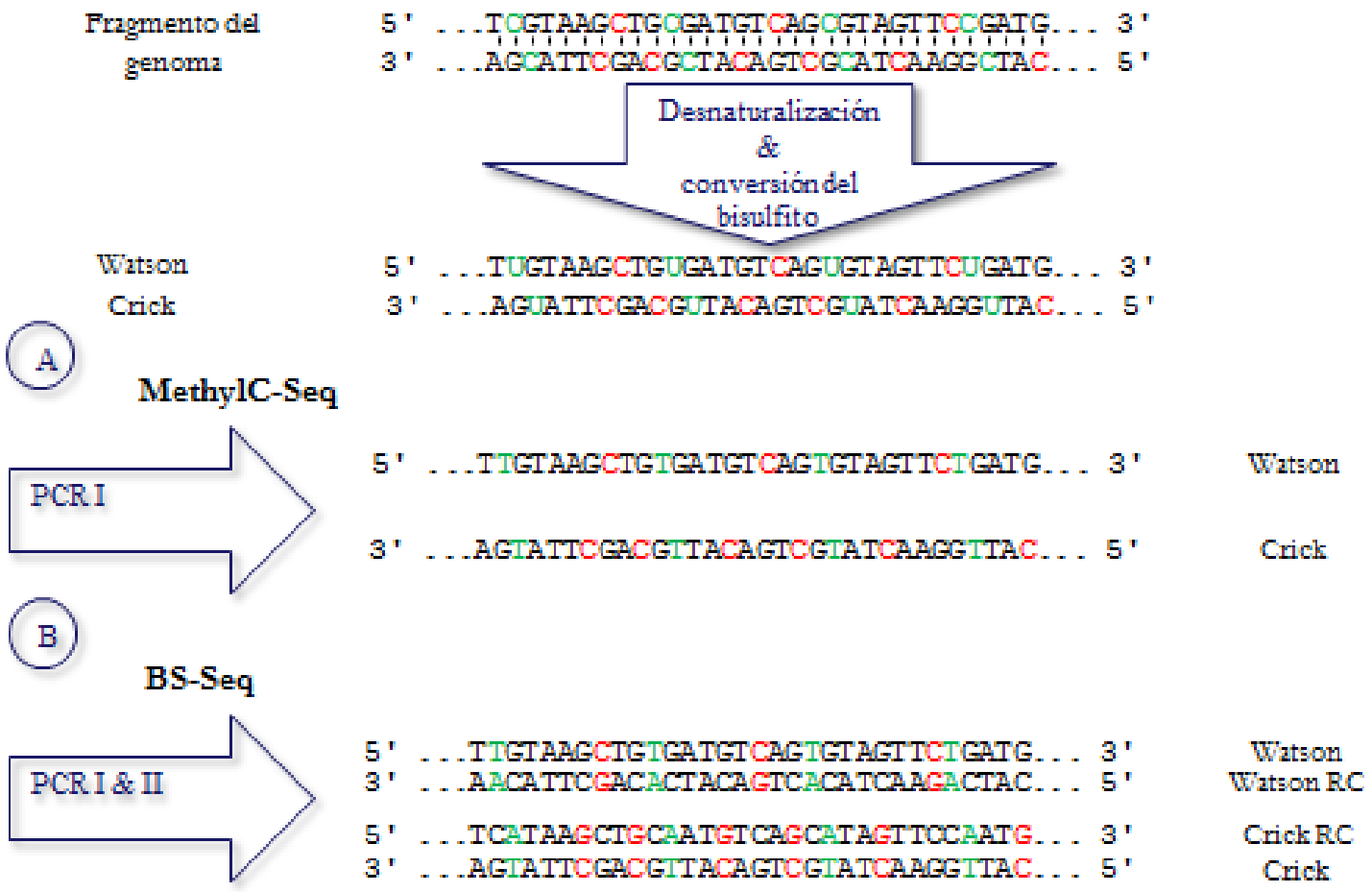
Ventajas

- Información de genoma completo
- Todos los contextos
- Valores específicos por citosina
- Información alelo específica

Inconvenientes

- Re-secuenciar genoma enteros
- Alineamiento de secuencias
- Errores de secuenciación y de conversión
- Variación de secuencia

Tratamiento con bisulfito



Reducción de la complejidad de secuencia en las regiones no metiladas
&
Aumentará el número de desemparejamientos durante el alineamiento

Alineamiento de lecturas tratadas con bisulfito

Opciones

1. **Probar todas las combinaciones posibles por lectura de C/T**
   - Inviable computacionalmente

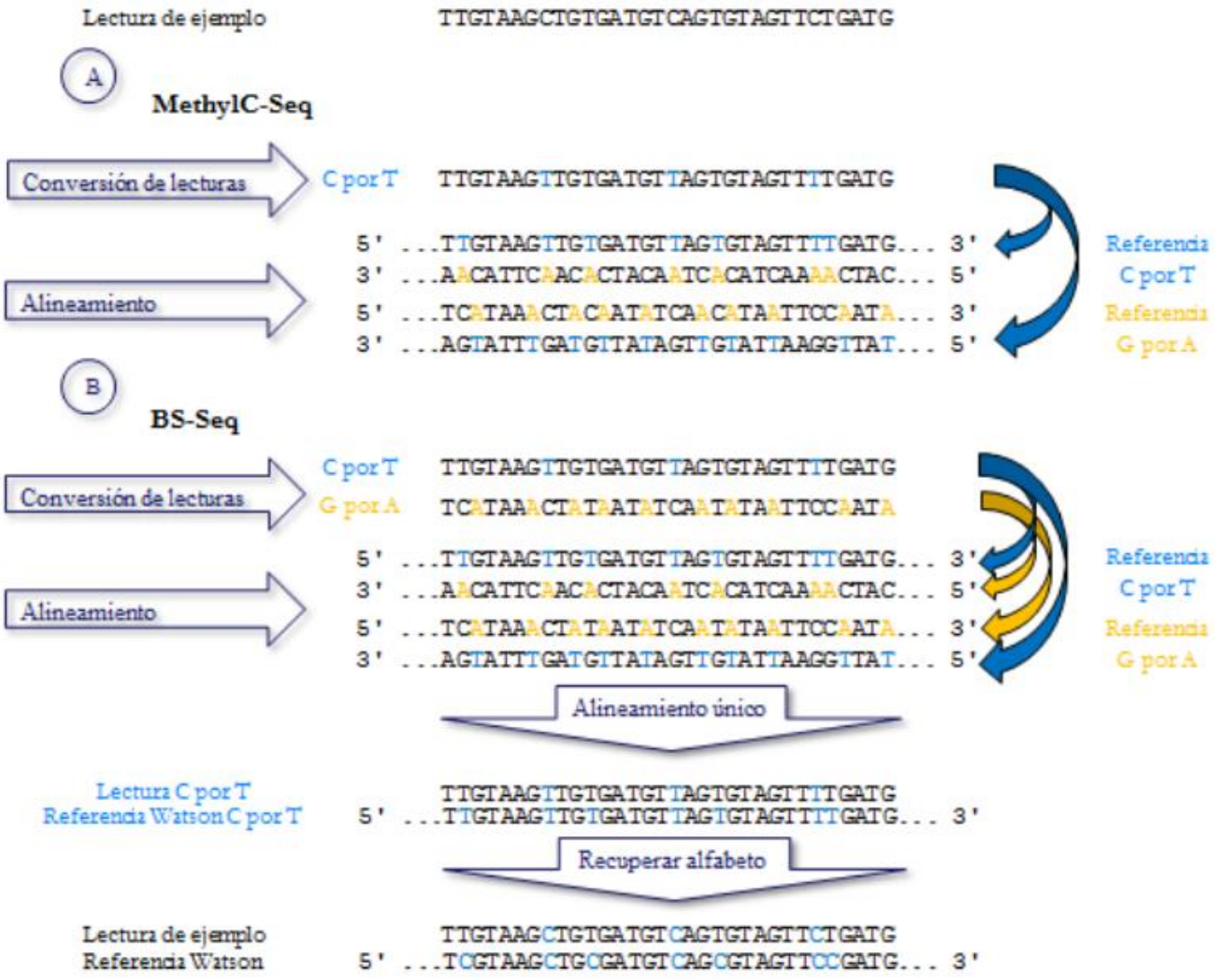
2. **Permitir un mayor número de desemparejamientos por lectura**
   - Mayor número de lecturas ambiguas
   - Mayor número de lecturas no alineadas
   - Tiempo de alineamiento incrementado

3. **Modificar la matriz de puntuación del alineamiento**
   - C referencia / T lectura > Emparejamiento
   - Mayor número de alineamientos en regiones metiladas
   - Sesgo en la estimación de la metilación

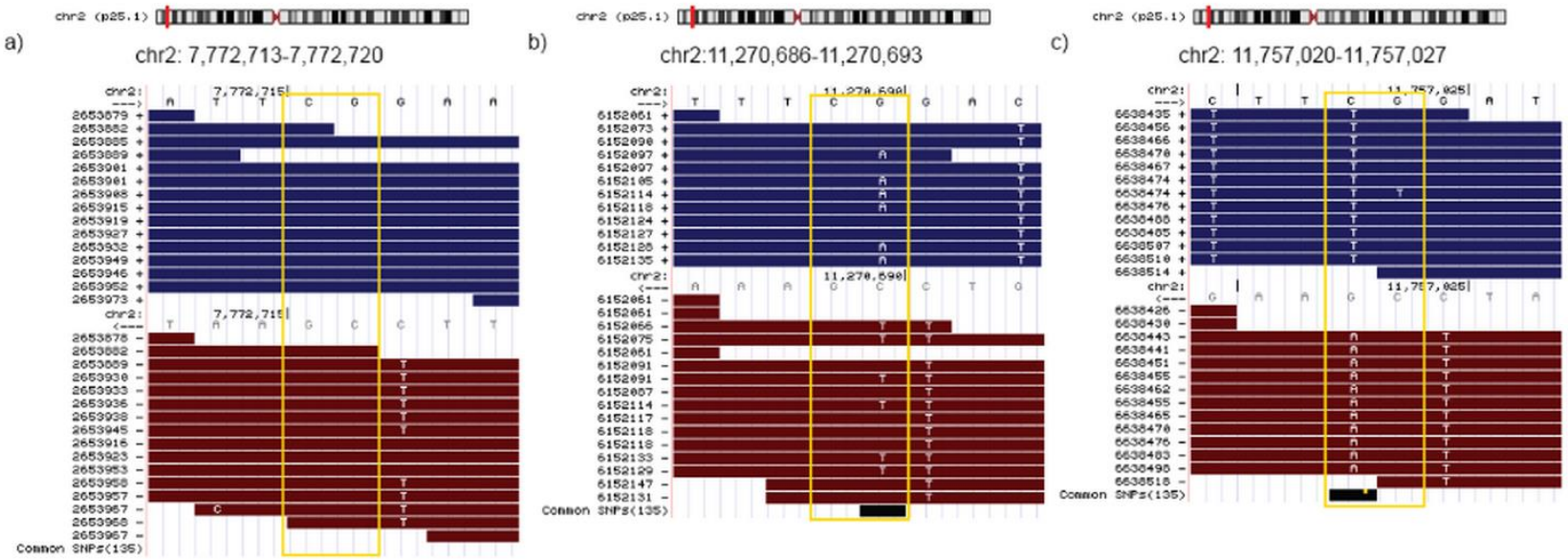
4. **Transformar el alfabeto de la referencia y de las lecturas a 3 letras**
   - Reducción de la complejidad de la secuencia
   - Aumento del espacio de búsqueda
   - Mayor número de lecturas ambiguas (no tanto como 2)

Alineamiento de lecturas tratadas con bisulfito

Alfabeto de 3 letras método de elección



Estimación de los niveles de metilación



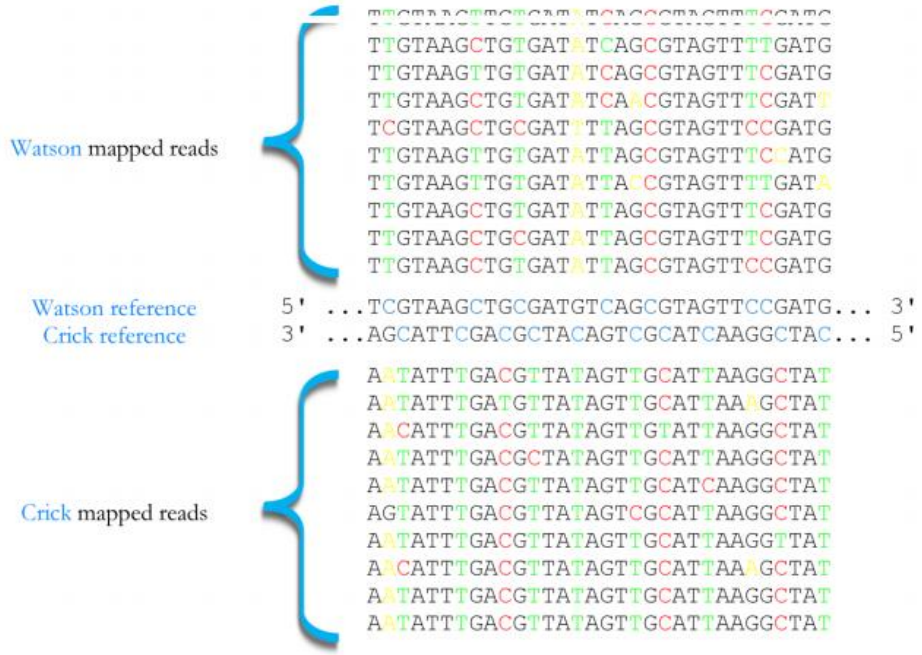
Homocigoto
CG / CG

Heterocigoto
CG / CA

Homocigoto
TG / TG

Variantes de secuencia pueden confundir los niveles de metilación estimados
C referencia / T lecturas > ¿Citosina no metilada o Variante?

Estimación de los niveles de metilación



Context	Start	Watson Level	Crick level	Merged Level	Watson SNV fraction	Crick SNV fraction	Result
CG (+/-)	2	0.1	0.2	0.15	0	0.9	rejected
CTT (-)	4	-	0	-	0	-	unMeth
CWG (+/-)	8	0.6	0.9	0.75	0	0	interMeth
CG (+/-)	11	0.2	0.1	0.15	0	0	unMeth
CAT (-)	13	-	0	0	1	-	rejected
CWG (+/-)	17	0.3	0.1	0.2	0.2	0	unMeth
CG (+/-)	20	1	0.9	0.95	0	0	meth
CTA (-)	22	-	0.1	-	0	-	unMeth
CCG (+)	27	0.2	-	-	-	0.2	unMeth
CG (+/-)	28	0.8	0.9	0.85	0.1	0	meth
CAT (-)	30	-	0	-	0.2	0	unMeth