

circRNA 全攻略手冊

I. 生物資訊預測 (此結果已完整輸出，不需再進行)

1. mapping reference : hg38 (我們用 hg 38。hg19 為較舊版本，可能也能參考)
2. mapping 策略 : BWA, 完成後存 SAM 檔
3. BAM 轉檔 : 將 SAM 轉為 BAM 檔
4. 利用 CIRI 進行 circRNA 預測 (pipeline檔案名稱 : [Jordan_bwa_CIRI.bash](#))

II. circRNA 位置註解與 primer 設計

1. circRNA 位置註解 (檔案名稱 : [circRNA candidates \(CIRI 預測\)\(最終\).xlsx](#))

Position	TID	GID	GeneSyn	ExonN	ExonM	Strand
chr10:118685983-118730410	ENST00000544392	ENSG00000151893	CACUL1	3	7	-
chr11:14771937-14789242	ENST00000282096	ENSG00000152270	PDE3B	3	4	+
chr11:17145668-17145942	ENST00000265970	ENSG00000011405	PIK3C2A	6	7	-
chr12:12908243-12909171	ENST00000014914	ENSG00000013588	GPRC5A	2	2	+
chr12:51055835-51056502	ENST00000262055	ENSG00000050426	LETMD	5	7	+
chr14:23033368-23033674	ENST00000361611	ENSG00000100804	PSMB5	2	2	-

* CIRI 結合蕭貴陽老師 Lab 的流程輸出的 excel 檔案概觀

Position :

CIRI 的輸出檔預測之 back splice 發生位置，例如: chr10:118685983-118730410。策略上主要依據這個位置取得 exon 序列的資訊 (注意 ! Position ID 在不同的人類基因體版本如 hg38 與 hg19 會有不同)。

TID :

此欄為不同分析流程 (蕭貴陽老師 Lab) 依據 CIRI 註解出的結果，為 “Transcripts (轉錄型)” 的 ID。

GID :

Ensembl 系統中給予的 “Gene” ID，很適合用來直接搜尋，再用 position 對。

ExonN及ExonM:

由蕭老師所註解的預測circRNA中的exon 組成。例如exonM =3, exonN=7 則表示該 circRNA 為某 transcript 中的 exon 組成 3>4>5>6>7 環狀。此部份僅供參考，我們知道真實的位置必須以 Position 為準。

以下為我整理過的完整候選circRNA 清單 (工作表名稱是 : \$P_value 雙過+任意>50)

Position	TID	GID	Predicted by NCHU Hisao's lab											Calculated by TMM normalization											Ensembl transcript link
			GeneSymbol	ExonN	ExonM	Strand	H1	H2	H3	N1	N2	N3	R1	R2	R3	FC N/H	FC R/H	P NH	P RH	P NR	H_mean	N_mean	R_mean		
chr10:7276892-7285954	ENST00000361972	ENSG00000198879	SFMBT2	5	7	-	93	105	123	60	51	67	69	58	58	0.709	0.731	0.010	0.007	0.667	107.000	59.333	65.000	http://www.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000361972	
chr10:7797047-7802854	ENST00000356708	ENSG00000165629	ATP5F1C	4	8	+	93	90	88	41	39	54	55	53	55	0.626	0.720	0.006	0.007	0.104	90.333	44.667	54.333	https://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000356708	
chr11:108148279-108161293	ENST00000065488	ENSG00000085491	SLC25A24	4	6	-	125	109	119	38	52	53	77	67	69	0.514	0.721	0.004	0.018	0.012	117.667	47.667	71.000	https://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000065488	
chr12:108652272-108654410	ENST00000261401	ENSG00000110880	CORO1C	7	8	-	62	78	79	41	36	47	49	45	55	0.724	0.819	0.008	0.034	0.151	73.000	41.333	49.667	https://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000261401	
chr12:42210355-42210680	ENST00000546726	ENSG00000015153	YAF2	3	4	-	106	104	116	37	46	65	85	64	76	0.574	0.826	0.007	0.038	0.047	108.667	49.333	75.000	https://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000546726	
chr14:211230319-21124229	ENST00000554969	ENSG00000092199	HNRNPC	5	6	-	52	48	58	51	59	54	56	49	59	1.328	1.244	0.022	0.024	0.386	52.667	54.667	54.667	https://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000554969	
chr14:65561337-65561766	ENST00000360689	ENSG00000033170	FUT8	3	3	+	124	104	115	161	150	194	141	133	128	1.866	1.402	0.002	0.012	0.009	114.333	168.333	134.000	https://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000360689	
chr15:67231814-67236820	ENST00000261880	ENSG00000103591	AAGAB	2	5	-	38	32	41	52	42	64	43	52	45	1.806	1.518	0.012	0.026	0.247	37.000	52.667	46.667	https://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000261880	
chr15:8012028-80122800	ENST00000261749	ENSG00000086666	ZFAND6	3	5	+	117	111	118	65	46	51	86	67	78	0.600	0.798	0.010	0.026	0.078	115.333	54.000	77.000	http://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000261749	
chr17:60265412-60294801	ENST00000300896	ENSG00000170832	USP32	6	9	-	61	66	69	38	24	25	51	40	38	0.572	0.790	0.014	0.029	0.145	65.333	29.000	43.000	http://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000300896	
chr19:40736986-40737769	ENST00000263370	ENSG00000086544	ITPKC	5	6	+	85	117	82	42	34	57	48	52	47	0.594	0.623	0.039	0.034	0.731	94.667	44.333	49.000	http://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000263370	
chr21:15014344-15043574	ENST00000400202	ENSG00000180530	NR1P1	2	2	-	324	371	365	278	321	350	347	330	321	1.142	1.132	0.043	0.018	0.867	353.333	316.333	332.667	http://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000400202	
chr2:72718103-72733118	ENST00000272427	ENSG00000144036	EXOC6B	3	5	-	144	181	161	68	66	76	97	92	99	0.552	0.713	0.001	0.005	0.000	162.000	70.000	96.000	http://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000272427	
chr3:119863419-119905852	ENST00000316626	ENSG00000082701	GSK3B	8	9	-	56	59	75	14	18	19	39	36	31	0.343	0.673	0.001	0.015	0.003	63.333	17.000	35.333	http://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000316626	
chr5:123545417-123557564	ENST00000345990	ENSG00000151292	CSNK1G3	3	4	+	127	149	166	72	65	100	106	101	100	0.683	0.838	0.017	0.023	0.093	147.333	79.000	102.333	http://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000345990	
chr5:145817894-145826200	ENST00000394450	ENSG00000186314	PRELID2	2	5	-	68	77	62	17	10	18	35	34	30	0.276	0.574	0.001	0.006	0.005	69.000	15.000	33.000	http://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000394450	
chr6:87215903-87218731	ENST00000369577	ENSG00000188994	ZNF292	2	4	+	99	82	86	26	17	29	51	42	36	0.341	0.576	0.003	0.017	0.029	89.000	24.000	43.000	https://asia.ensembl.org/Homo_sapiens/Transcript/View?transcript=ENST00000369577	
																-0.5		<0.05	<0.05	<0.05	>0.05	<50	<50	<50	

<接著是利用 e!Ensembl 網站進行序列的取得>：網址：https://www.ensembl.org/Homo_sapiens/Info/Index

雖然我已經把這些序在的連結放到 [circRNA candidates \(CIRI 預測\)\(最終\).xlsx](#) 裡的 **Ensembl transcript link** 了

- (1) 我們直接以 TID 搜尋 (以 *circFMBT2* 為例)，輸入其 ID “ENST00000361972” 搜尋 (請記得選擇物種為人類)，畫面如下：

e!Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog

Human (GRCh38.p12) ▼

Search Human (*Homo sapiens*)

Search all categories ▼ ENST00000361972 **Go**

e.g. [BRCA2](#) or [17:63992802-64038237](#) or [rs699](#) or [osteoarthritis](#)

Genome assembly: GRCh38.p12 (GCA_000001405.27)

- More information and statistics
- Download DNA sequence (FASTA)
- Convert your data to GRCh38 coordinates
- Display your data in Ensembl

Other assemblies

GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart ▼ **Go**

View karyotype

Example region

- (2) 可以看到我們搜尋到的地方被標注了黃色背景，點選第一列的 “SFMBT2-201 (Human Transcript)”。

← → ↻ 🏠 https://www.ensembl.org/Human/Search/Results?q=ENST00000361972;facet_feature_type=;site=ensembl;fac

e!Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog

New Search

Current selection:
< all Species
Only searching Human

Restrict category to:
Gene 1
Transcript 1
Protein Family 1

Per page:
10 25 50 100

Layout:
Standard Table

Tip:
Help and Documentation can be searched from the homepage! Just type in a term you want to know more about, like non-synonymous SNP.

Only searching Human ▼ ENST00000361972 **Q**

3 results match ENST00000361972 when restricted to **species: Human**

SFMBT2-201 (Human Transcript)
ENST00000361972 10:7158624-7409341:-1
Scm like with four mbt domains 2 [Source:HGNC Symbol;Acc:HGNC:20256]
Q5VUG0.1 (UniProtKB/Swiss-Prot) is used as supporting evidence for transcript **ENST00000361972**
Location • External Refs. • cDNA seq. • Exons • Variant table • Protein seq. • Population
• Protein summary

PTHR12247_SF62 (Human Protein Family)
PTHR12247_SF62
Ensembl protein family PTHR12247_SF62 [SCM WITH FOUR MBT DOMAINS 2]: 1 gene / 4 proteins in homo sapiens.

SFMBT2 (Human Gene)
ENSG00000198879 10:7158624-7411486:-1
Scm like with four mbt domains 2 [Source:HGNC Symbol;Acc:HGNC:20256]
CCDS31138.1 (CCDS) is an external reference matched to Transcript **ENST00000361972**
Variant table • Phenotypes • Location • External Refs. • Regulation • Orthologues • Gene tree

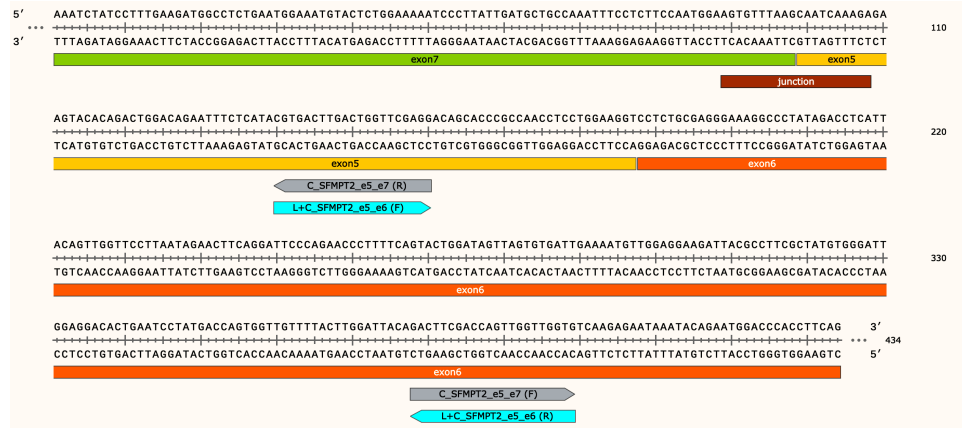
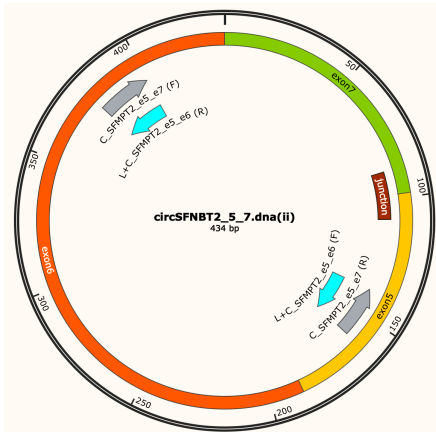
<< < 1 > >>

Ensembl release 96 - April 2019 © [EMBL-EBI](#)
[Permanent link](#) - [View in archive site](#)

2. primer 設計策略 (以circSFNB2 為例)

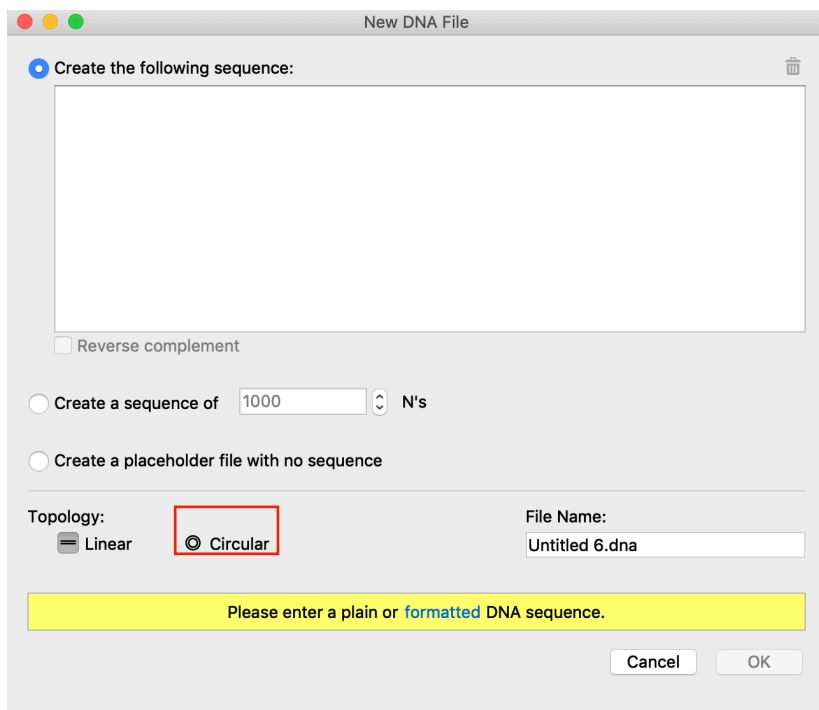
- Junction 上的 primer (代號: C (circ)):

如圖中的灰色 primer，目標在於 exon 7 和 exon 5 中間的 junction，預期夾出部分的 exon 5, 6 和完整的 exon 7。



Tip1. 在把這三段 exon 貼到 SnapGene 之前可以依 exon 7, exon 5, exon 6 的順序貼上，這樣在設計夾 junction 的 primer 時比較好像畫面。

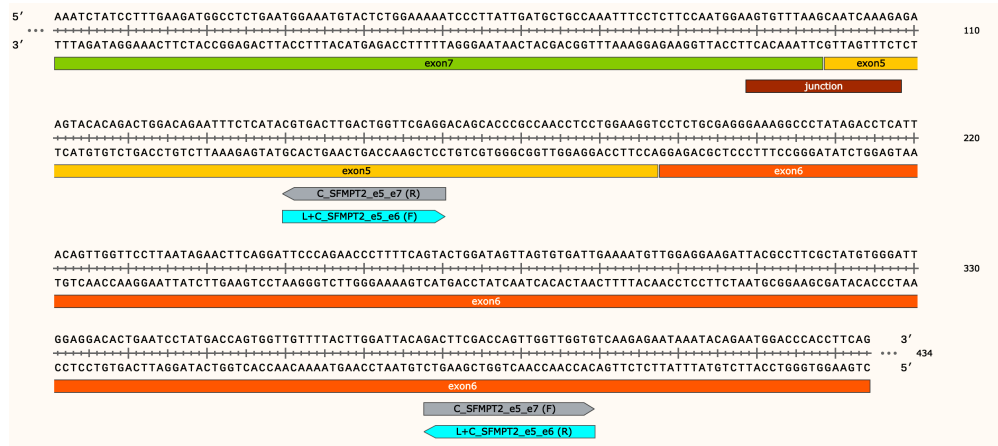
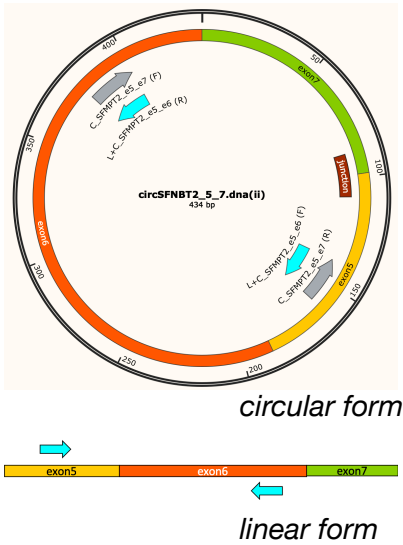
Tip2. 序列貼到 SnapGene 的時候可以直接“topology”選項中勾選“Circular”，這樣在 map 圖上即可方便的視覺化。



Tip3. Junction 的長度以 120-160 bp 來設計 primer 為佳，不過以一般 PCR 驗證而言，200-400 bp 亦可接受。

• CircRNA 與 linear RNA 都可以 P 出的 primer (代號 : L+C (linear + circ)):

如圖中的藍色 primer，目標在於 exon 5 和 exon 6。

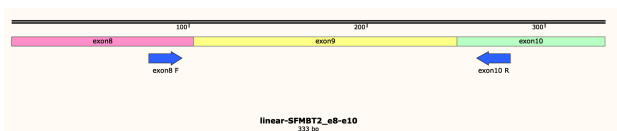


Tip1. 可以發現設計這類型的 primer 可以直接拿專門 P junction 的 primer 位置來設計，因為只是 primer 夾的”方向”不同而已。

Tip2. 這種類型的 primer 意義在於它可以在相同的 exons 序列上同時夾出 linear 和 circular form 的片段，很適合用於比較。

• Linear RNA 專一的 primer (代號 : L (linear)):

在相同的原始基因中，找一段確定只有線性產物的 exon 組合 (避開可能形成 circRNA 的 exon 區域)，用它來代表線性的 RNA 產物。



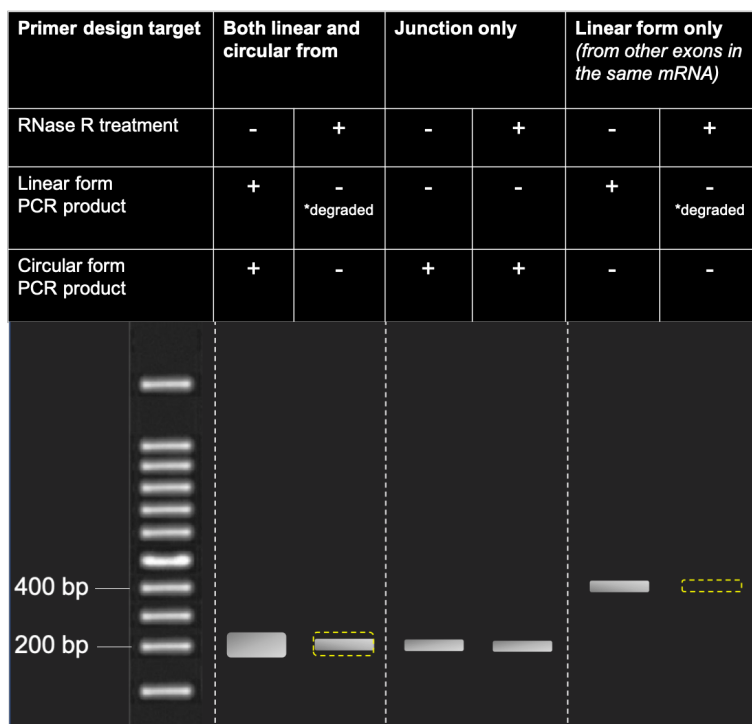
linear only: exon 8 - exon 10

(避開了 exon 5 - exon 7 的可能環型區域)

Tip1. 這段長度建議設計接近於 circular form 能夾出的長度。

<預期跑膠狀況視覺化>

- 假設：1. 假設對 circRNA 上的 junction 預期 P 出的長度約為 200 bp。
2. 假設對 linear form 和 circular form 上共享的片段預期 P 出的長度也約為 200 bp。
3. 假設在避開可能形成 circRNA 的 exons 區域後，能找到位在同 mRNA 上，但不同 exons 的區域有一段 exons 組合，可以作為 linear form 代表的 mRNA，而該段 exons 組合長度約 400 bp。



III. circRNA 抽取與 RNase R treatment

- 抽取 circRNA 的方式如同實驗室常規的 NucleoZOL 抽取方法。
- 抽取完成後的 total RNA 習慣配製成 200 ng/μl，以方便後續實驗。
- 將實驗分成兩組：RNase R treatment & Non-RNase R treatment。我們預期 RNase R treatment 組別只會留下 circular form 的 RNA。

RNase R treatment 經典條件如下：

Reagent	Volume (μl)	Final unit	Company / 代理商
RNase R (20U/μl)	0.5	10 U	Lucigen® / 波仕特
10x RNase R buffer	1	1x	Lucigen® / 波仕特
RNA template (200 ng/μl)	5	1000 ng	
Nuclease-free water	3.5		

- RNase R treatment 組在加完上述試劑後置入 37°C 培養箱 (或放 PCR 內設定 37°C program)，反應 30 - 60 分鐘 (經典時間為 30 分鐘，此條件可以依實驗情形調整反應時間)。

*註：RNase R 單價 14,000。請小心使用，平均一個反應 tube 就會花 700 元。

*註：冰太久可能導致酵素活性下降 / 新買的酵素活性相對高，必須先做好適當條件

IV. 反轉錄成 cDNA

1. RT-PCR 反應試劑

Reagent	Volume (μ l)	Company
10x RT random primer	2	Applied Biosystems®
10 x RT buffer	2	Applied Biosystems®
25x dNTP mix (100 mM)	0.8	Applied Biosystems®
MultiScribe Reverse Transcriptase (50 U/ μ l)	1	
Nuclease-free water	4.2 (RNase R treatment) 9.2 (Non-RNase R treatment)	
RNA template	使總量達 1000 ng:	
• RNase R treatment 組： 反應完後濃度為 100 ng/ μ l	10 (RNase R treatment)	
• Non-RNase R treatment 組： 原管濃度固定為 200 ng/ μ l	5 (Non-RNase R treatment)	
<i>*Total (in a PCR tube)</i>	<i>*20</i>	

2. PCR 條件

Temperature ($^{\circ}$ C)	Time (min)	cycle
25	10	1
37	60	1
37	60	1
85	5	1
15	∞	1

V. PCR 偵測 circRNA junction 與 linear form RNA 之存在

1. PCR 反應試劑 (PCR tube 內總體積：25 or 50 μ l 皆可)

Reagent	Volume (μ l)	Final unit	Company
2x PCR Master Mix	12.5	1x	Thermo Scientific®
F'+R' primer (5 μ M)	4	1 μ M	自行設計
cDNA template (轉完cDNA濃度剩 50 ng/ μ l)	2	100 ng	
Nuclease-free water	6.5		

2. PCR 條件

Temperature (°C)	Time	cycle
95	3 min	1
95	30 s	25-35 (自行依狀況判斷)
Tm°C - 5 (自行依梯度PCR決定)	30 s	
72	1 min	想看得很清楚：35 想做定量：25-30
72	10 min	1

* 註：適當提升溫度有助於解決 off-target 的問題。

VI. DNA 電泳驗證

1. 配膠 (尺梳可自行選擇)：

- 大片膠：60 ml TAE buffer + 2% agarose (適用於 0.1 - 3 kb target DNA) + 6 μ l SYBR® safe DNA stain
- 小片膠：30 ml TAE buffer + 2% agarose (適用於 0.1 - 3 kb target DNA) + 3 μ l SYBR® safe DNA stain

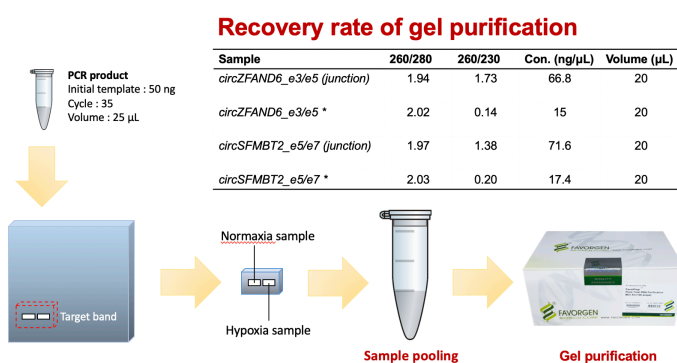
2. Loading (PCR Master Mix 產品為一鍋式製程，PCR 後不需要再添加 loading dye)：

- Sample loading (5 μ l for 最小尺梳)
- Marker loading (本實驗選用 100 bp 型，因產物大小約落在 200-500 bp。3 μ l for 最小尺梳)

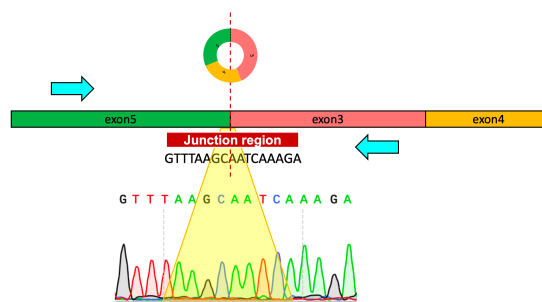
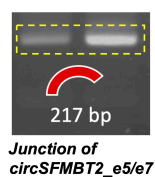
3. 100V 開始電泳，當 marker 與 sample dye (黃色) 的指示達倒數第二條橫線時停止 (約 30 min)。

VII. 純化膠體將感興趣的 junction region 送定序

1. 在 UV 光下將條帶切割下來放入 eppendorf。
2. 利用 FAVORGEN® 的 kit 進行 gel proliferation。

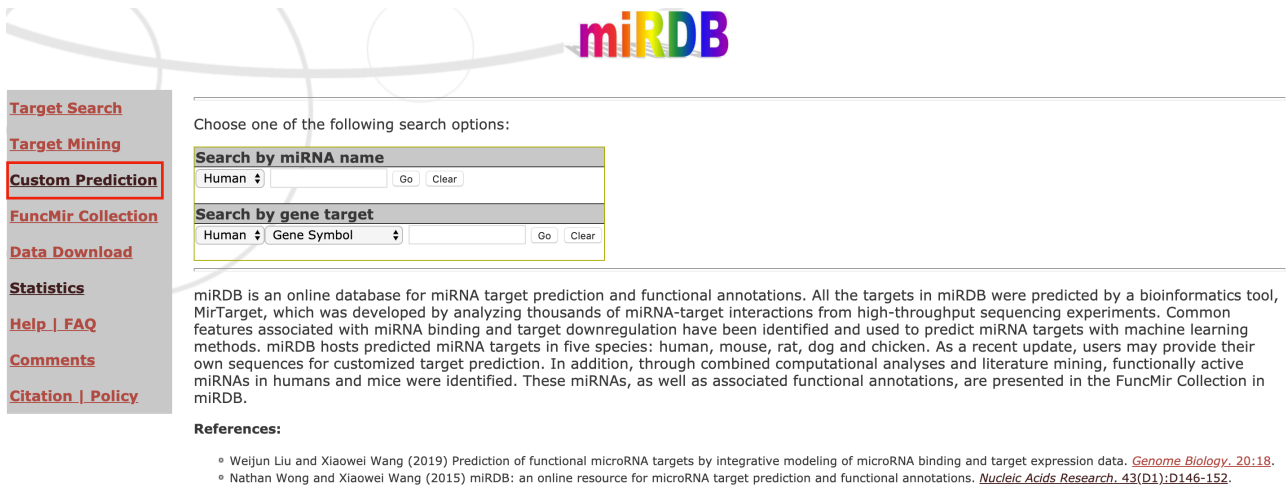


3. 送定序 (總量需答 500 ng)



VIII. 都驗證到了，就來看看 microRNA sponge 吧

1. 使用 miRDB 進行預測 (網址：<http://mirdb.org/>)，選擇 **Custom Prediction**



The screenshot shows the miRDB website interface. On the left is a navigation menu with options: Target Search, Target Mining, Custom Prediction (highlighted with a red box), FuncMir Collection, Data Download, Statistics, Help | FAQ, Comments, and Citation | Policy. The main content area has a heading "miRDB" and a sub-heading "Choose one of the following search options:". Below this are two search forms: "Search by miRNA name" and "Search by gene target". The "Search by miRNA name" form has a dropdown menu set to "Human" and a "Go" button. The "Search by gene target" form has a dropdown menu set to "Human" and a "Gene Symbol" dropdown, with "Go" and "Clear" buttons. Below the search forms is a paragraph of text describing miRDB as an online database for miRNA target prediction and functional annotations, mentioning the MirTarget tool and the species it supports (human, mouse, rat, dog, and chicken). At the bottom, there is a "References:" section with two citations: Weijun Liu and Xiaowei Wang (2019) and Nathan Wong and Xiaowei Wang (2015).

2. 我們的目的是送 circRNA 的序列，讓他去預測可能有的 miRNA。所以 Submission type 要勾選 “mRNA target sequence”。



The screenshot shows the miRDB website interface for a custom miRNA target sequence search. The navigation menu is on the left, with "Custom Prediction" highlighted. The main content area has a heading "miRDB" and a sub-heading "miRDB provides additional web interface below for more flexible miRNA target search. You may search targets by providing your own sequence. In addition, target search can also be performed for unconventional sites in the coding region or 5'-UTR." Below this is a section titled "1) Target prediction with user-provided miRNA or gene target sequence". It contains a form with a "Species:" dropdown set to "Human" and a "miRNA Sequence" input field. The "Submission type" dropdown is set to "miRNA target Sequence" (highlighted with a red box). Below the dropdown are two radio buttons: "Example miRNA submission" (unchecked) and "Example target mRNA submission" (checked). There is a large text area for the sequence and "Go" and "Clear" buttons at the bottom. At the bottom of the page, there is a note: "Custom miRNA sequences must be between 17 and 30 bases long; custom mRNA target sequences must be between 100 and 30,000 bases long. Please submit only one sequence at a time, consisting of A, T or U, C and G. Please note that all spaces, line breaks, and numbers within the submission will be removed, and multiple lines will be joined together to create a single sequence."

3. 剩下就是看書出的結果了。這些預測中的 seed region 都是完美符合才會顯示出來，即使分數只有50。