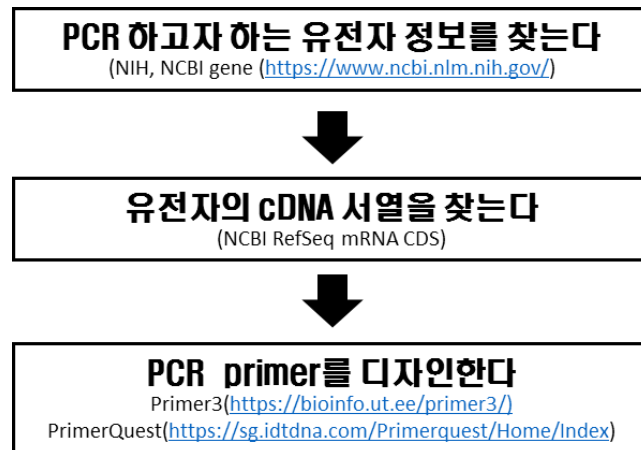


생명 과학자 기초 체력 다지기 <제2회> PCR

이제욱 (오송첨단의료산업진흥재단)

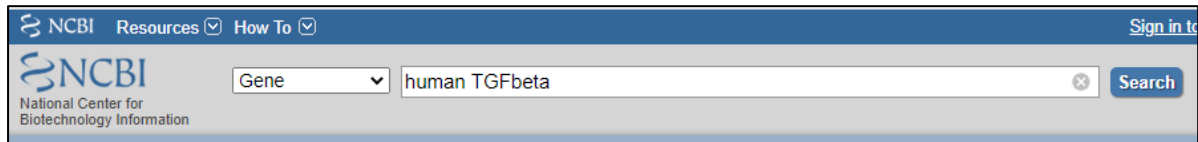
<실례: 유전자 서열 정보로부터 PCR Primer 디자인 하기>

인간의 TGFbeta1 유전자 전체를 증폭하기 위한 PCR primer를 디자인하는 방법을 알아 보자.



[유전정보로부터 PCR primer 디자인 흐름도]

먼저, 인간의 TGFbeta 유전자 정보를 NCBI로부터 찾는다. 미국 국립보건원 (NIH)의 **NCBI** (National Center for Biotechnology Information) (<https://www.ncbi.nlm.nih.gov/>)에 접속한 후 DB를 "Gene"으로 선택하고, "human TGFbeta"를 입력하고 "Search"를 클릭한다.



다음 검색 결과 중 “첫번째 유전자명 (TGFB1)”을 클릭한다.

Search results

Items: 1 to 20 of 833 << First < Prev Page of 42 Next > Last >>

i Showing Current items.

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> TGFB1 ID: 7040	transforming growth factor beta 1 [<i>Homo sapiens</i> (human)]	Chromosome 19, NC_000019.10 (41330323..41353922, complement)	CED, DPD1, IBDIMDE, LAP, TGF-beta1, TGFB, TGFbeta	190180
<input type="checkbox"/> HMGA2 ID: 8091	high mobility group AT-hook 2 [<i>Homo sapiens</i> (human)]	Chromosome 12, NC_000012.12 (65824460..65966291)	BABL, HMGI-C, HMGIC, LIPO, SRS5, STQL9	600698

검색된 결과를 scroll down하다가 우측에 gene="TGFB1"가 있는 좌측의 “CDS”를 클릭한다.

```

CDS      join(5883..6237,10472..10632,14063..14180,16680..16757,
        16897..17044,26646..26799,27717..27875)
        /gene="TGFB1"
        /gene_synonym="CED; DPD1; IBDIMDE; LAP; TGF-beta1; TGFB;
        TGFbeta"
    
```

CDS를 클릭하여 나온 결과를 계속 scroll down하여 **NCBI Reference Sequences (RefSeq), mRNA and Proteins(s)** 아래의 “**NM_000660.7**”를 클릭한다. 참고로 NM은 mRNA, 즉 cDNA에 대한 정보이고, NP는 protein에 대한 정보를 나타낸다.

NCBI Reference Sequences (RefSeq)

RefSeqs maintained independently of Annotated Genomes

These reference sequences exist independently of genome builds. [Explain](#)

Genomic

1. **NG_013364.1 RefSeqGene**

Range	5005..28604
Download	GenBank , FASTA , Sequence Viewer (Graphics)

mRNA and Protein(s)

1. [NM_000660.7](#) → [NP_000651.3](#) transforming growth factor beta-1 proprotein preproprotein
[See identical proteins and their annotated locations for NP_000651.3](#)

Status: REVIEWED

검색된 결과를 scroll down하다가 우측에 gene="TGFB1"가 있는 좌측의 "CDS"를 클릭한다.

```
CDS      879..2051
         /gene="TGFB1"
         /gene_synonym="CED; DPD1; IBDIMDE; LAP; TGF-beta1; TGFB;
         TGFBeta"
```

그럼 다음과 같이 "highlight"된 DNA 서열이 나온다. "Highlight된 부분은 human TGFB1 full length cDNA 서열"을 나타낸다. Highlight된 서열의 처음에서 커서를 놓고 클릭한 채로 마지막 서열까지 선택한 후 "Ctrl+C" 혹은 "우클릭하여 copy를 선택"하여 서열을 복사한 후, MS Word 또는 한글에서 "Ctrl+V"로 서열을 붙인다.

```
781 ttttgccggg agatccccag cccctgcagg ggccggggcct ccccaccaca ccagccctgt
841 tcgcgctctc ggccagtgcc gggggcgccg cctccccat gccgccctcc gggctgccc
901 tgctgccgct gctgctaccg ctgctgtggc tactggtgct gacgcctggc cggccggccg
961 cgggactatc cacctgcaag actatcgaca tggagctggt gaagcgggaag cgcctcgagg
1021 ccatccgcgg ccagatcctg tccaagctgc ggctcgccag ccccccgagc cagggggagg
1081 tgccgcccgg cccgctgcc gaggccgtgc tcgccctgta caacagcacc cgcgaccggg
1141 tggccgggga gagtgcagaa ccggagcccg agcctgaggc gcactactac gccaaaggagg
1201 tcacccgcgt gctaattggtg gaaaccaca acgaaatcta tgacaagttc aagcagagta
1261 cacacagcat atatatgttc ttcaacacat cagagctccg agaagcggta cctgaaccgg
1321 tgttgctctc ccgggcagag ctgctgtgc tgaggctcaa gttaaaagtg gagcagcacg
1381 tggagctgta ccagaaatac agcaacaatt cctggcgata cctcagcaac cggctgctgg
1441 caccagcga ctgccagag tggttatctt ttgatgtcac cggagttgtg cggcagtggt
1501 tgagccgtgg aggggaaatt gagggcttc gccttaggc ccaactgctc tgtgacagca
1561 gggataaac actgcaagtg gacatcaac ggttcactac cggccgccga ggtgacctgg
1621 ccaccattca tggcatgaac cggcctttc tgcttctcat ggccaccccg ctggagaggg
1681 ccagcatct gcaaagctc cggcaccgc gagccctgga caccaactat tgcttcagct
1741 ccacggagaa gaactgctg gtgcggcagc tgtacattga cttccgcaag gacctggct
1801 ggaagtggat ccacgagcc aagggtacc atgccaactt ctgcctcggg cctgcccct
1861 acatttggag cctggacacg cagtacagca aggtcctggc cctgtacaac cagcataacc
1921 cggcgccctc ggccggcggc tgctgctgc cgcaggcgt ggagccgctg cccatcgtgt
1981 actacgtggg ccgcaagccc aagggtgagc agctgtccaa catgatcgtg cgctcctgca
2041 agtgcagctg aggtcccgcc ccgccccgcc ccgccccggc aggcccgcc ccacccgcc
2101 ccgccccgc tgccttgccc atgggggctg tatttaagga caccgtgcc ccaagcccac
```

유전자 이름을 다음과 같은 형태 (>hTGFB1)로 붙인다. 참고로, (>) 기호 다음에 유전자 이름을 붙이는 것은 유전자 서열과 파일명을 구분하는 약속이며, 이러한 형식을 "FASTA format"이라 한다.

>hTGFb1

```
at gccgcctcc gggctgcggc
901  tgctgcccgt  getgctaccg  ctgctgtggc  tactggtgct  gacgcctggc  cggccggccg
961  cgggactatc  cacctgcaag  actatcgaca  tggagctggt  gaagcggaaq  cgcacgcagg
1021 ccatccgcgg  ccagatcctg  tccaagctgc  ggctcgccag  cccccgagc  cagggggagg
1081 tgccgcccgg  cccgctgccc  gaggccgtgc  tcgccctgta  caacagcacc  cgcgaccggg
1141 tggccgggga  gagtgcagaa  cgggagcccg  agcctgaggc  cgactactac  gccaaaggag
1201 tcacccgctg  gctaatgggt  gaaacccaca  acgaaatcta  tgacaagtgc  aagcagagta
1261 cacacagcat  atatatgttc  ttcaacacat  cagagctccg  agaagcggta  cctgaaaccg
1321 tgttgctctc  ccgggcagag  ctgctgtgct  tgaggctcaa  gttaaaagtg  gagcagcacg
1381 tggagctgta  ccagaaatac  agcaacaatt  cctggcgata  cctcagcaac  cggctgtggg
1441 caccagcga  ctgcccagag  tggttatctt  ttgatgtcac  cggagtgtg  cggcagtggt
1501 tgagccgtgg  aggggaaatt  gagggttct  gccttagcgc  ccactgctcc  tgtgacagca
1561 ggataaacac  actgcaagtg  gacatcaacg  ggttcactac  cggccgcccga  ggtgacctgg
1621 ccaccattca  tggcatgaac  cggccttcc  tgcttctcat  ggccaccccg  ctggagaggg
1681 ccagcatct  gcaaagctcc  cggcaccgcc  gagccctgga  caccaactat  tgcttcagct
1741 ccacggagaa  gaactgctgc  gtgcggcagc  tgtacattga  cttccgcaag  gacctcggct
1801 ggaagtggat  ccacgagccc  aagggtacc  atgccaaact  ctgcctcggg  cctgcccct
1861 acatttgag  cctggacacg  cagtacagca  aggtcctggc  cctgtacaac  cagcataacc
1921 cgggcgectc  gggcgcccg  tgctgctgct  cgcagggcgt  ggagccgctg  cccatcgtgt
1981 actacgtggg  cgcgaagccc  aagggtggag  agctgtccaa  catgatcgtg  cgctcctgca
2041 agtgcagctg  a
```

여기까지가 증폭하고자 하는 유전자 서열을 찾는 방법이다. 첫번째 3 bases인 “atg”가 “Start codon”이고, 마지막 3 bases인 “tga”가 “Stop codon”이다.

그럼 앞서 찾은 유전자 서열을 이용하여 “PCR primer”를 디자인해 보자.

유전자 서열 중 숫자는 무시되므로 신경 쓰지 않아도 된다. 그래도 신경 쓰인다면, **NCBI ORFfinder** (<https://www.ncbi.nlm.nih.gov/orffinder/>)에서 제거할 수 있다. ORFfinder 주소를 클릭하면 “Enter Query Sequence”라는 박스가 나온다. 그 박스에 “human TGFb1 서열을 이름부터 선택하여 copy한 후 붙여 넣고” 그 아래 “Submit”을 클릭한다.

NCBI Resources How To Sign in to NCBI

ORFfinder PubMed Search

COVID-19 is an emerging, rapidly evolving situation.
 Get the latest public health information from CDC: <https://www.coronavirus.gov>.
 Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
 Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

Open Reading Frame Finder

ORF finder searches for open reading frames (ORFs) in the DNA sequence you enter. The program returns the range of each ORF, along with its protein translation. Use ORF finder to search newly sequenced DNA for potential protein encoding segments, verify predicted protein using newly developed SMART BLAST or regular BLASTP.

This web version of the ORF finder is limited to the subrange of the query sequence up to 50 kb long. Stand-alone version, which doesn't have query sequence length limitation, is available for [Linux x64](#).

Examples (click to set values, then click Submit button):

- NC_011604 Salmonella enterica plasmid pWES-1; genetic code: 11; 'ATG' and alternative initiation codons; minimal ORF length: 300 nt
- NM_000059; start codon: 'ATG only'; minimal ORF length: 150 nt

Enter Query Sequence

Enter accession number, gi, or nucleotide sequence in FASTA format:

```
>hTGFb1
at gccgcctccc gggctgccc
  901 tgcctccgct gctgctaccg ctgctgtggc tactggtgct gacgcttggc cggccggccg
  961 cgggactatc cactgcaag actatcgaca tggagctggg gaagcgggag cgcctcgagg
 1021 ccatccggcg ccagatcctg tccaagctgc ggctgcccag cccccgagc cagggggagg
 1081 tggccggcgg cccgctgccc gaggccgtgc tcgcccgtga caacagcacc cgcgaccggg
 1141 tggccgggga gaggcagaaa cgggagcccg agcctgaggc cgactactac gccaaaggag
 1201 tcaccggcgt gtaaatgggt gaaaccacac agaaatcta tgacaagttc aagcagagta
 1261 cacacagcat atatatgttc ttcaacacat cagagctccg agaagcggta cctgaaccgg
 1321 tgtgtctctc ccgggcagag ctgctctctc tgaggctcaa gttaaaagtg gacgagcag
```

Start Search / Clear

Submit Clear

검색 결과 중 가장 긴 ORF1을 선택하고, ORF1,4,6이 있는 박스 위의 Protein FASTA의 메뉴를 “CDS FASTA”로 바꾸고, **Download marked set**을 클릭하면 “해당 유전자 서열을 저장할까요” 라는 화면이 뜬다. 그럼 원하는 곳에 “서열을 저장”한다.

Open Reading Frame Viewer Help

Sequence

ORFs found: 7 Genetic code: 1 Start codon: 'ATG' only

ORF1 0 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1 K 1,050 1,100 1,173

ORFfinder_9.25.234138898

ORF7 ORF2 ORF1 ORF6 ORF3 ORF4 ORF5

1: 1..1.2K (1,173 nt) Tracks shown: 2/5

Six-frame translation...

ORF1 (390 aa) Display ORF as... Unmark

```
>|c1|ORF1
MPPSGRLRLLPLLLPLLLVLTGPRPAAGLSTCKTIDMELVKKRRIEAIK
GQLLSKLRLASPPSQEVPVPGPLPEAVLALYNSTRDVRVAGESAEPEPEPE
ADYYAKVTRVLNVEITHIELYDKFKQSTHSYIFNITSELRLEAVPEPVL
SRAELRLLRLLKLVQHVLYQKYSNHSRHYLSNRLAPSDSEPHLSFDV
TGVNRQMLSRGGEIEGFRLSAHCSDSRDNTLQVDINGFTTGRGRLATZ
HGMNRPFLLLMATPLERAQLQSSRRRALDNYCFSSTKEMCCVQRQLYZ
DFRXDLGHWKIHPEPKYHANFCLGCPYINSLDTQYSKVLALYNQHNPQA
SAAPCCVQALEPLTVIYVYVGRKPKVEQLSNMIVRSCCKC
```

Label	Strand	Frame	Start	Stop	Length (nt aa)
ORF1	+	1	<1	1173	1173 390
ORF4	-	1	1038	586	453 150
ORF6	-	2	749	360	390 129
ORF3	+	2	953	1147	195 64

Mark subset... Marked: 1 Download marked set as CDS FASTA

그럼 다음과 같이 서열이 저장된다. 유전자 이름을 ">hTGfb1"으로 변경한다.

```
>hTGfb1
ATGCCGCCCTCCGGGCTGCGGCTGCTGCCGCTGCTGCTACCGCTGCTGTGGCTACTGGTGTGACGCCTGCC
GGCCGGCCGGCCGGGACTATCCACCTGCAAGACTATCGACATGGAGCTGGTGAAGCGGAAGCGCATCGAGGCC
ATCCGGCGCCAGATCCTGTCCAAGCTGCGGCTCGCCAGCCCCCGAGCCAGGGGGAGGTGCCGCCGGCCC
GCTGCCCGAGGCCGTGCTCGCCCTGTACAACAGCACCCGCGACCGGGTGGCCGGGGAGAGTGAGAAACCGG
AGCCCGAGCCTGAGGCCGACTACTACGCCAAGGAGGTCACCCGCGTCTAATGTTGGTGAAGAACCCACAACGAAAT
CTATGACAAGTTCAAGCAGAGTACACAGCATATATATGTTCTTCAACACATCAGAGCTCCGAGAAGCGGTACCT
GAACCCGTGTTGCTCTCCGGGCGAGGCTGCGTCTGCTGAGGCTCAAGTAAAAGTGGAGCAGCAGCGTGGAG
CTGTACCAGAAATACAGCAACAATTCCTGGCGATACCTCAGCAACCGGCTGCTGGCACCAGCGACTCGCCAGA
```

그럼 유전자 서열을 사용하여 "PCR prime"를 디자인해 보자.

Primer 디자인은 각자 취향에 따라 상업용 프로그램을 사용하거나 무료 웹사이트 혹은 primer 합성 회사의 웹사이트에 제공하는 디자인 tool 을 사용하면 된다. 가장 널리 사용하는 무료 웹사이트는 **Primer3** (<https://bioinfo.ut.ee/primer3/>)이다. 그 외 **PrimerQuest** (<https://sg.idtdna.com/Primerquest/Home/Index>)도 사용하기 편리하지만, 회원 등록 후 사용이 가능하다.

여기서는 "PrimerQuest Tool"을 사용하여 Sequence Entry 에 앞서 찾아 둔 서열만 복사하여 붙여 넣는다.

The screenshot shows the 'PrimerQuest Tool' interface. At the top, there are buttons for 'ASSAY DESIGN' and 'RESULTS'. On the right, there are buttons for 'HELP' and 'ABOUT'. The main section is titled 'Sequence Entry' and has a dropdown menu set to 'Enter sequence(s) manually'. Below this is a text box containing the DNA sequence from the previous image. To the right of the text box is a 'Hints' section with the following instructions:

- Paste your sequence into the Textbox
- Add up to 50 sequences in FASTA format
- Sequence length must be greater than 80 bases

위의 페이지를 scroll down 하면 맨 아래에 primer 시작과 종료 서열을 지정할 수 있다. **전체 서열이 "1173 bp (Stop codon 포함)"**이므로 Forward primer의 시작 서열을 "1", Reverse primer의 시작 서열을 "1173"으로 입력하고 "GET ASSAYS"를 클릭한다.

Primer Force Location Settings

Force Forward Start Position (5' End)

Force Forward Stop Position (3' End)

Force Reverse Start Position (5' End)

Force Reverse Stop Position (3' End)

[GET ASSAYS](#)

그럼 다음과 같이 default 로 5 개의 primer sets 을 보여 준다.

<input type="checkbox"/> Set 1 hTGFb1						
Amplicon Length: 1173 View Assay Details	Forward	Start	Stop	Length	Tm	GC%
		1	18	17	71	76.5
	Reverse	1153	1174	21	69	61.9
<input type="checkbox"/> Set 2 hTGFb1						
Amplicon Length: 1173 View Assay Details	Forward	1	18	17	71	76.5
	Reverse	1152	1174	22	71	63.6
<input type="checkbox"/> Set 3 hTGFb1						
Amplicon Length: 1173 View Assay Details	Forward	1	19	18	72	77.8
	Reverse	1153	1174	21	69	61.9
<input type="checkbox"/> Set 4 hTGFb1						
Amplicon Length: 1173 View Assay Details	Forward	1	19	18	72	77.8
	Reverse	1152	1174	22	71	63.6
<input type="checkbox"/> Set 5 hTGFb1						
Amplicon Length: 1173 View Assay Details	Forward	1	18	17	71	76.5
	Reverse	1154	1174	20	67	60

첫번째 primer 의 "view assay details"을 클릭하면 다음과 같은 정보를 보여 준다.

PrimerQuest Tool

ASSAY DESIGN
RESULTS
HELP
ABOUT

hTGFb1 Assay Set 1 Details

BACK TO RESULTS

Parameter Set: General PCR (Primers only)
 Sequence Name: hTGFb1
 Amplicon Length: 1173

		Start	Stop	Length	Tm	GC%
Forward	ATGCCGCCCTCCGGGCT (Sense)	1	18	17	71	76.5
Reverse	TCAGCTGCACTGCAGGAGCG (AntiSense)	1153	1174	21	69	61.9

Base	Sequence
1	ATGCCGCCCTCCGGGCTGCGGGTGGTGGCGGTGCTGCTACCGCTGCTGTGGCTACTGCTGCTGACGGCTGGCCGGGGGGGGGGGGGAGCTATCCACCTGCA
101	AGACTATCGACATGGAGCTGGTGAAGCGGAAGCGCATCGAGGCCATCCGGGGCCAGATCCTGTCCAGCTGCGGGTGGCCAGCCCCCGAGCCAGGGGGGA
201	GGTGGCGCCGGGCGCGCTGCCCCGAGGCGGTGCTCGCCCTGTACAAACGACCCCGGACCGGGTGGCCGGGGAGAGTGCAGAACCGGAGCCCGAGCCTGAG
301	GCCGACTACTACGCCAAGGAGGTCAACCGCGTGTAAATGGTGAACCCACAAACGAAATCTATGACAACTTCAAGCAGAGTACACACAGCATATATATGT
401	TCTTCAACACATCAGAGCTCCGAGAAGCGGTACCTGAAACCCGCTGTCTCTCCGGGCGAGAGCTGCGCTGCTGAGGCTCAAGTAAAGTGGAGCAGCA
501	CGTGGAGCTGTACAGAAATACAGCAACAATTCCTGGCGATACCTCAGCAACCGGCTGCTGGCACCCAGGACTCGCCAGAGTGTATCTTTTGATGTC
601	ACGGAGTGTGGCGGCTGGTGGCGGTGGAGGGGAAATGAGGGCTTTCGCCCTAGGGCCCACTGCTCCTGTGACAGCAGGGATAACACACTGCAAG
701	TGACATCAACGGSTTCACTACCGGCGCCGAGGTGACCTGGCCACCAATTCATGSCATGAACCGGCTTTCTGCTTCTCATGGCCACCCCGCTGGAGAG
801	GGCCACGATCTGCAAAAGCTCCCGGACCCGCGAGCCCTGGACACCAACTATTGCTTCACTCCAGGAGAAAGAACTGCTGCTGGCGGAGCTGTACATT
901	GACTTCGCAAGGACCTCGGGTGGAAAGTGGATCCAGAGCCCAAGGGTACCATGCCAACTTCTGCTCGGGGCGCTGCCCTACATTGGAGCCTGGACA
1001	CGCAGTACAGCAAGGTCTCGCCCTGTACAAACAGCATAAACCGGGCGCCTCGGGCGGCGCTGCTGCTGCGCCAGGCGCTGGAGCCGCTGCCATCTG
1101	GTACTACGTGGCGCAAGCCCAAGGTGGAGCAGCTTCCAAATGATCGTGGCTCCTGCAAGTGCAGCTGA

그 다음으로, "qPCR primer"를 디자인해 보자.

동일한 tool 을 사용하여 Custom Design Parameters 를 "qPCR"로 선택하고, "Get Assays"를 클릭하면, qPCR primer 와 probe 를 찾아준다. 두 번째가 "TaqMan 방식"이고, 세 번째가 "일반적인 Dye qPCR 방식"이다.

Custom Design Parameters

Task Settings

Results

Results to Return (1-50)

Vary Forward Primer Location
Off
Low
Med
High

Vary Reverse Primer Location
Off
Low
Med
High

Set Design Parameters for...

- General PCR (Primers Only)
- qPCR (2 Primers + Probe)
- qPCR Intercalating Dyes (Primers only)
- Forward Only - Sequencing
- Reverse Only - Sequencing

PrimerQuest Tool

ASSAY DESIGN RESULTS HELP ABOUT

Assay Set Locations for hTGFb1

ADD TO ORDER DOWNLOAD ASSAYS CUSTOMIZE ASSAY DESIGN SHOW CUSTOM TARGET REGIONS

Set 1 hTGFb1

Amplicon Length: 112 [View Assay Details](#)

	Start	Stop	Length	Tm	GC%
Forward	501	523	22	62	50
Probe	527	551	24	68	50
Reverse	593	613	20	62	50

Set 2 hTGFb1

Amplicon Length: 83 [View Assay Details](#)

	Start	Stop	Length	Tm	GC%
Forward	641	660	19	62	52.6
Probe	675	699	24	68	50
Reverse	704	724	20	62	50

Set 3 hTGFb1

Amplicon Length: 112 [View Assay Details](#)

	Start	Stop	Length	Tm	GC%
Forward	593	613	20	62	50
Probe	638	662	24	67	50
Reverse	684	705	21	62	47.6

해당 primer 의 “view assay details”을 클릭하면, 다음과 같은 정보를 보여 준다.

PrimerQuest Tool

ASSAY DESIGN RESULTS HELP ABOUT

hTGFb1 Assay Set 1 Details

[BACK TO RESULTS](#)

Parameter Set: RT-qPCR (Primers with Probe)
 Sequence Name: hTGFb1
 Amplicon Length: 112

	Start	Stop	Length	Tm	GC%
Forward	501	523	22	62	50
Probe	527	551	24	68	50
Reverse	593	613	20	62	50

Base	Sequence
1	ATGCCGCOCTCCGGCTCGGGCTGCTGCCGCTGCTGCTACCGCTGCTGTGGCTACTGGTGTGACGCOCTGGCCGGCCGGCCCGGGGACTATCCACCTGCA
101	AGACTATCGACATGGAGCTGGTGAAGCGGARGGCGCATCGRGGCCATCCGCGGCCAGATCCTGTCCARGCTGCGGCTCGCCAGCCOCCGAGCCAGGGGA
201	GGTGCCOCCOCCGGCTGCCCCGAGGCCGTGCTCGCCCTGTACAAACAGCACCCCGGACCCGGGTGGCCGGGGAGAGTGCAGAACCCGGAGCCCGAGCCTGAG
301	GCCGACTACTACGCCAAGGAGGTCAOCCCGCTGCTAATGGTGGAAACCCACAAACGAAATCTATGACAAATTCCAGCAGAGTACACACAGCATATATATGT
401	TCTTCAACACATCAGAGCTCCGAGAAGCCGTACCTGAACCCGCTGCTCTCCCGGGCAGAGCTGCGTCTGCTGAGGCTCAAGTTAAAAGTGGAGCAGCA
501	CGTGGAGCTGTACCGAAATACAGCAACAATTCCTGGCGATACCTCAGCA ACCCGCTGCTGGCCOCCAGCGACTCGCCAGRGTGGTTATCTT TTGATGTC
601	ACCCGAGTTGTG CGGCAGTGGTTGAGCCGTGGAGGGGAAATAGAGGGCTTTTCGCTTAGCCGCCACTGCTCCTGTGACAGCAGGGATAACACACTGCAAG
701	TGGACATCAACGGGTTCACTACCGGCCCGGAGGTGACCTGGCCACCATTCATGGCATGAACCGGCTTTTCCTGCTTCTCATGGCCACCCCGCTGGAGAG
801	GGCCCCAGCATCTGCAAGCTCCCGGCACCCCGAGCCCTGGACACCAACTATTGCTTCAGCTCCACGGGAGAAGACTGCTGCGTGGCGCAGCTGTACATT
901	GACTTCCGCAAGGACCTCGGCTGGAAAGTGGATCCACGAGCCOCCAGGGCTACCATGCCAATCTTGCCTCGGGCCCTGCCCTACATTTGGAGCCTGGACA
1001	CGCAGTACAGCAAGTCTCTGGCCCTGTACAAOCCAGCATAAACCAGGGCCCTCGCCGGCCCGTGTGCTGCGTGGCCCGAGGGCCGTGGAGCCGCTGCCCATCGT
1101	GTACTACGTGGGCCAGGCCAAGGTGGAGCAGCTGTCCAACATGATCGTGGCTCCTGCAAGTGCAGCTGA