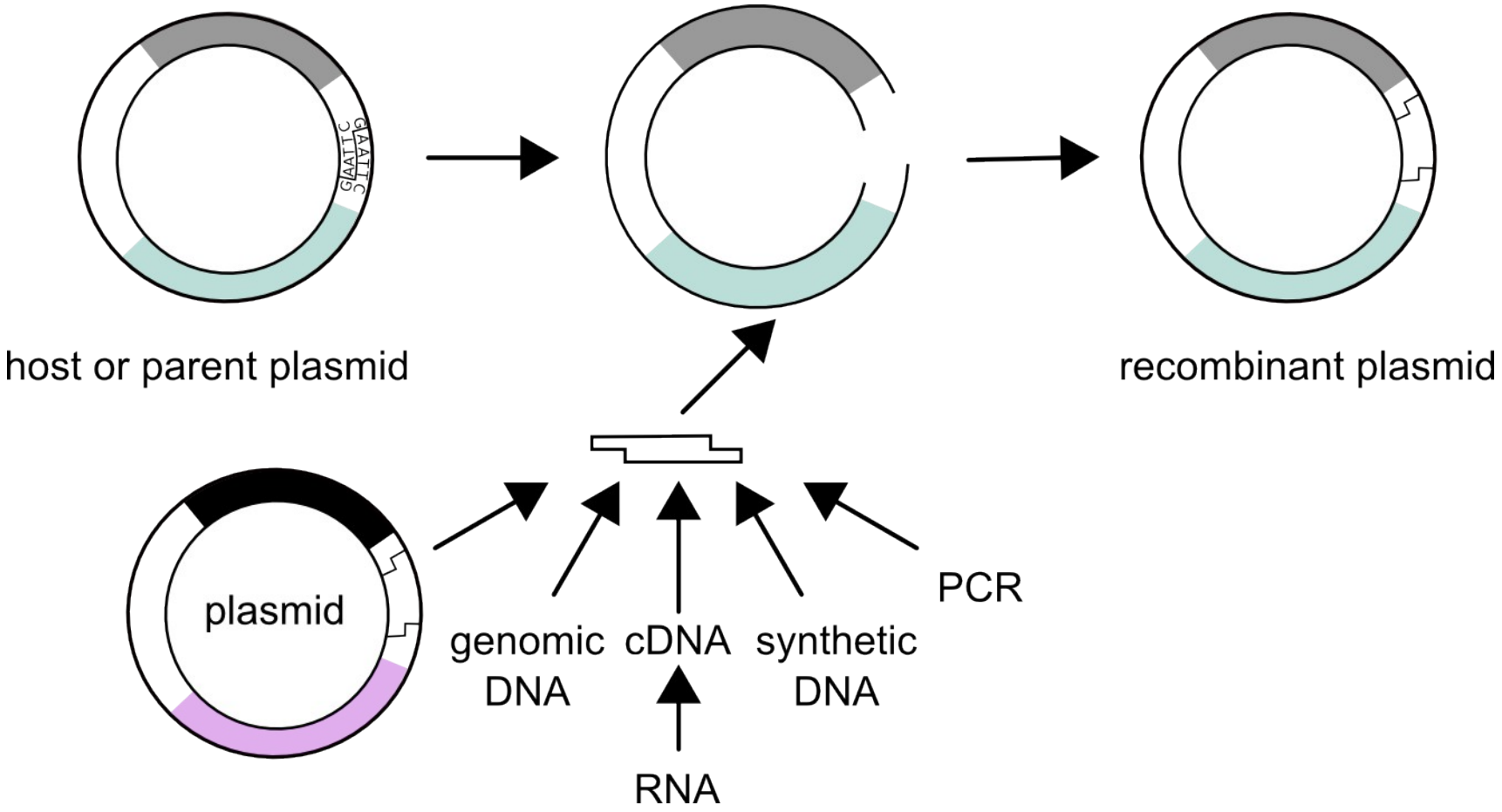
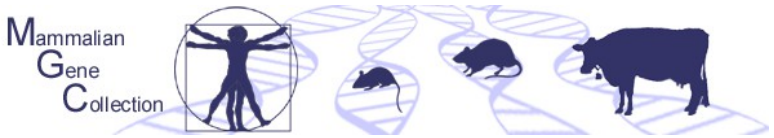
Sources of Sequences



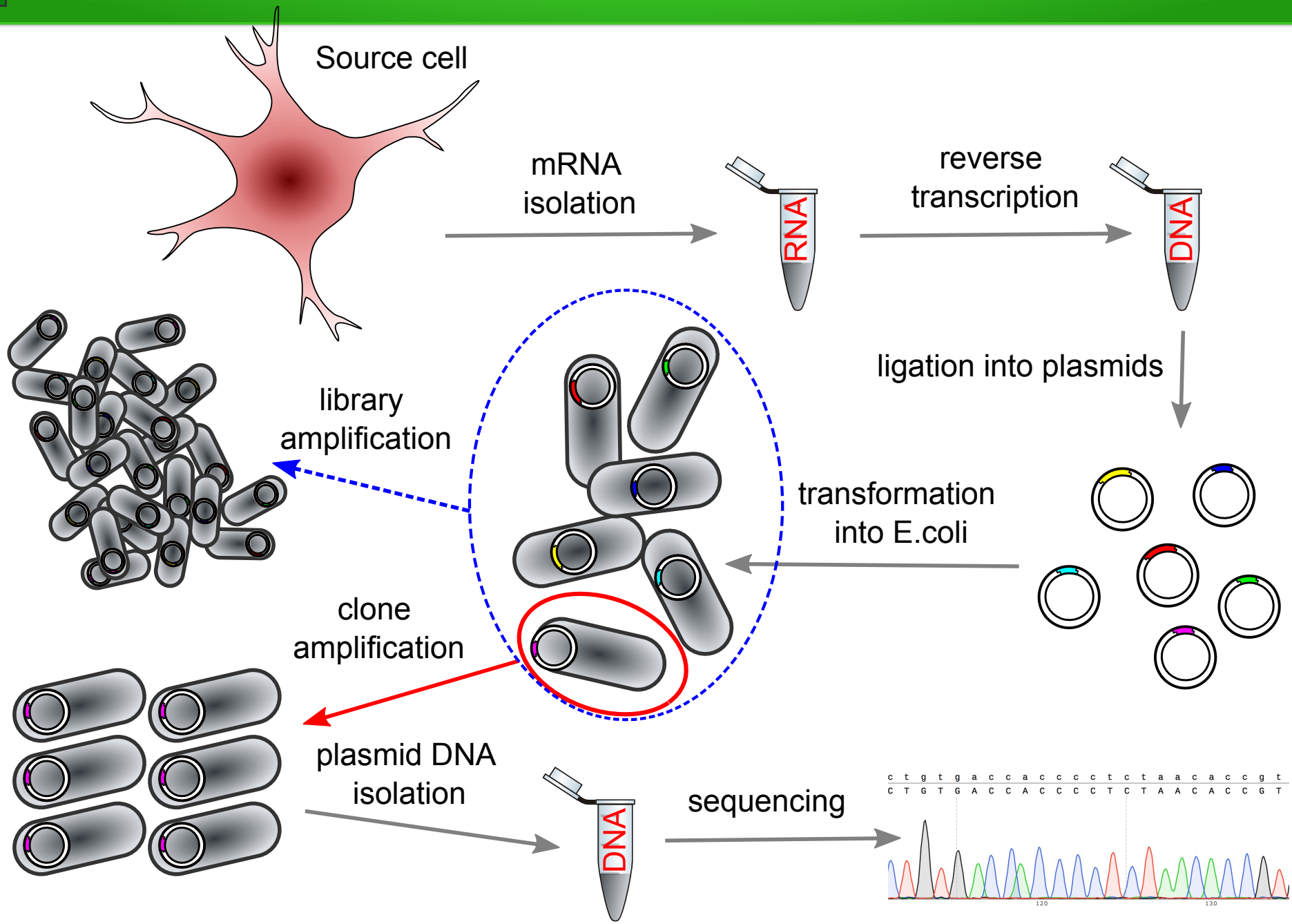
Gene Collections



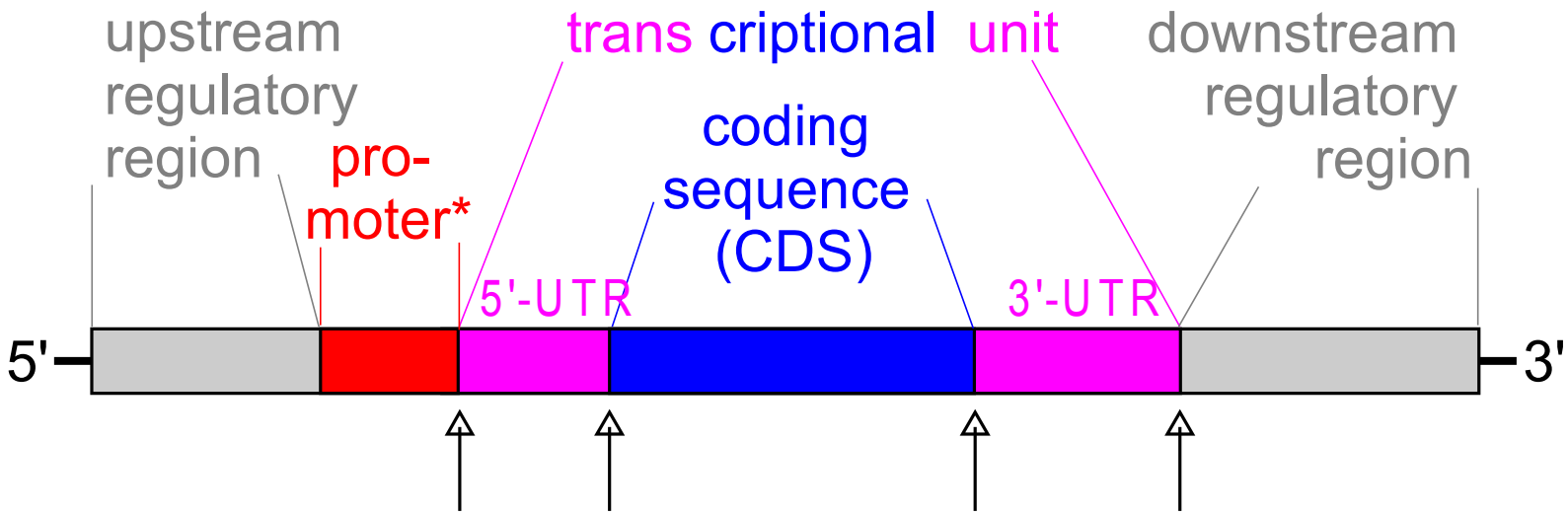
- MGC (Mammalian Gene Collection = human, mouse, rat, bovine): http://mgc.nci.nih.gov/ (most of these clones are available for a nominal fee from http://www.biocenter.helsinki.fi/bi/gbu/orf/)

  A guide for finding and evaluating MGC clones: http://mgc.nci.nih.gov/files/GuideToFindingEvaluating_MGC_Clones.pdf
- Zebrafish gene collection: http://zgc.nci.nih.gov/
- Xenopus gene collection: http://xgc.nci.nih.gov/
- MGC, honey bee, chicken, fugu, Drosophila, C. elegans: http://www.lifesciences.sourcebioscience.com/
- RIKEN (http://dna.brc.riken.jp/en/resource150en.html)
- Long cDNAs (~4.5kb):
  Kazusa (http://www.kazusa.or.jp/huge/, KIAA designation)
  NEDO (http://www.kazusa.or.jp/NEDO, FLJ designation)

How to make a cDNA library



Reverse transcription into cDNA



- House keeping genes and small ORFs are overrepresented in cDNA libraries
- Priming: “radom hexamers”, ”oligo(dT)-primer”, gene-specific primers
- cDNA library vs. cDNA expression library
- ORF clones (no 5'- & 3'-UTRs), with sequence-specific primers
- cDNA libraries→ cDNA collections
- Reverse transcriptase (error rate: 1:17,000-1:30,000, max. length ~15kb)

Gateway site-specific recombination cloning system

Based on λ phage integration system:

attB1xattP1→attL1
attB2xattP2→attL2

attL1xattR1→attB1
attL2xattR2→attB2

+/- stop codon

8 aa between GOI and tag (from att site)

ccdB containing plasmids must be propagated in a gyrA462+ strain!

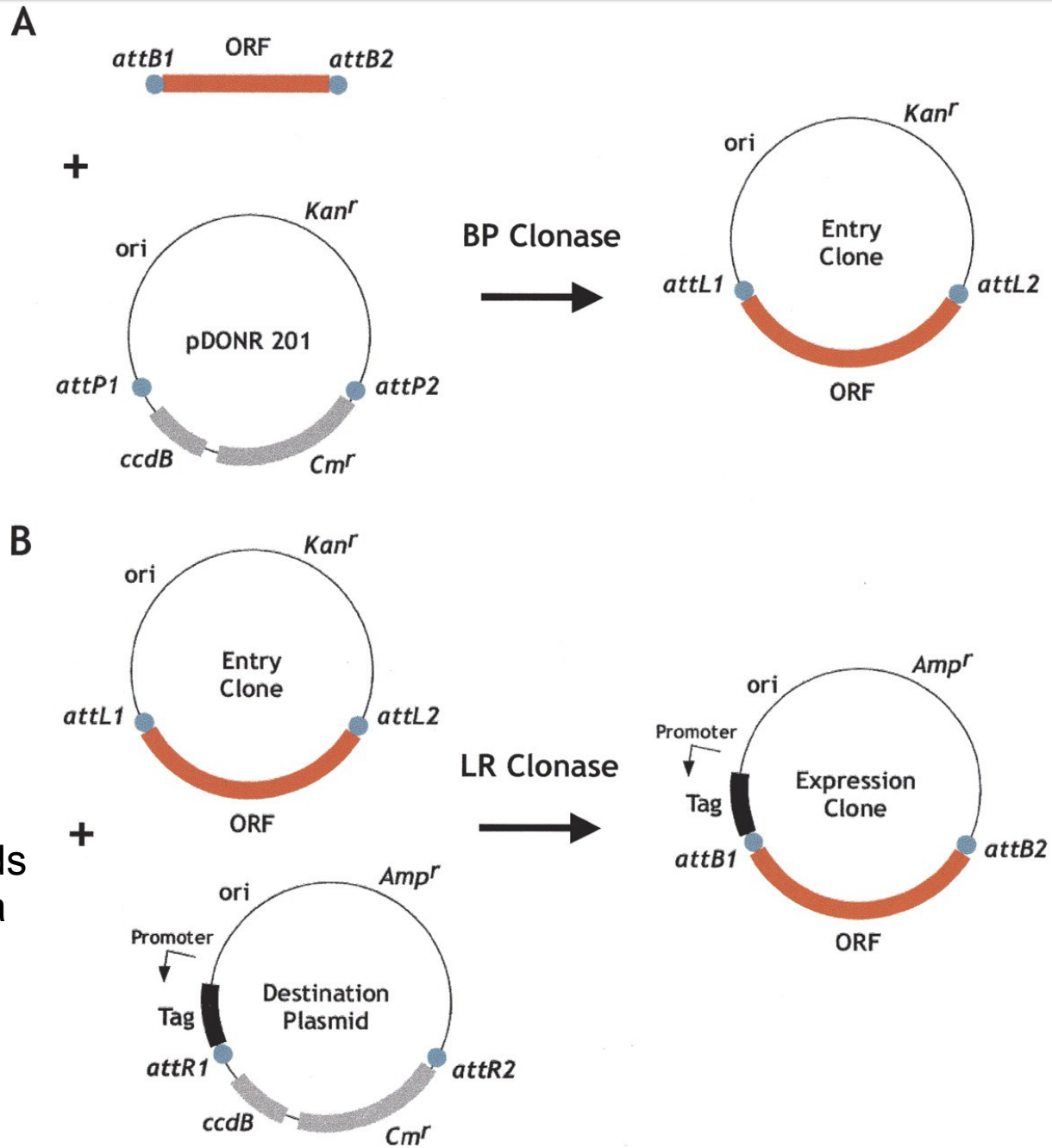


Figure from:
Marsischky G & LaBaer J Genome Res. 2004; 14: 2020-2028

In-Fusion/Creator Cloning System

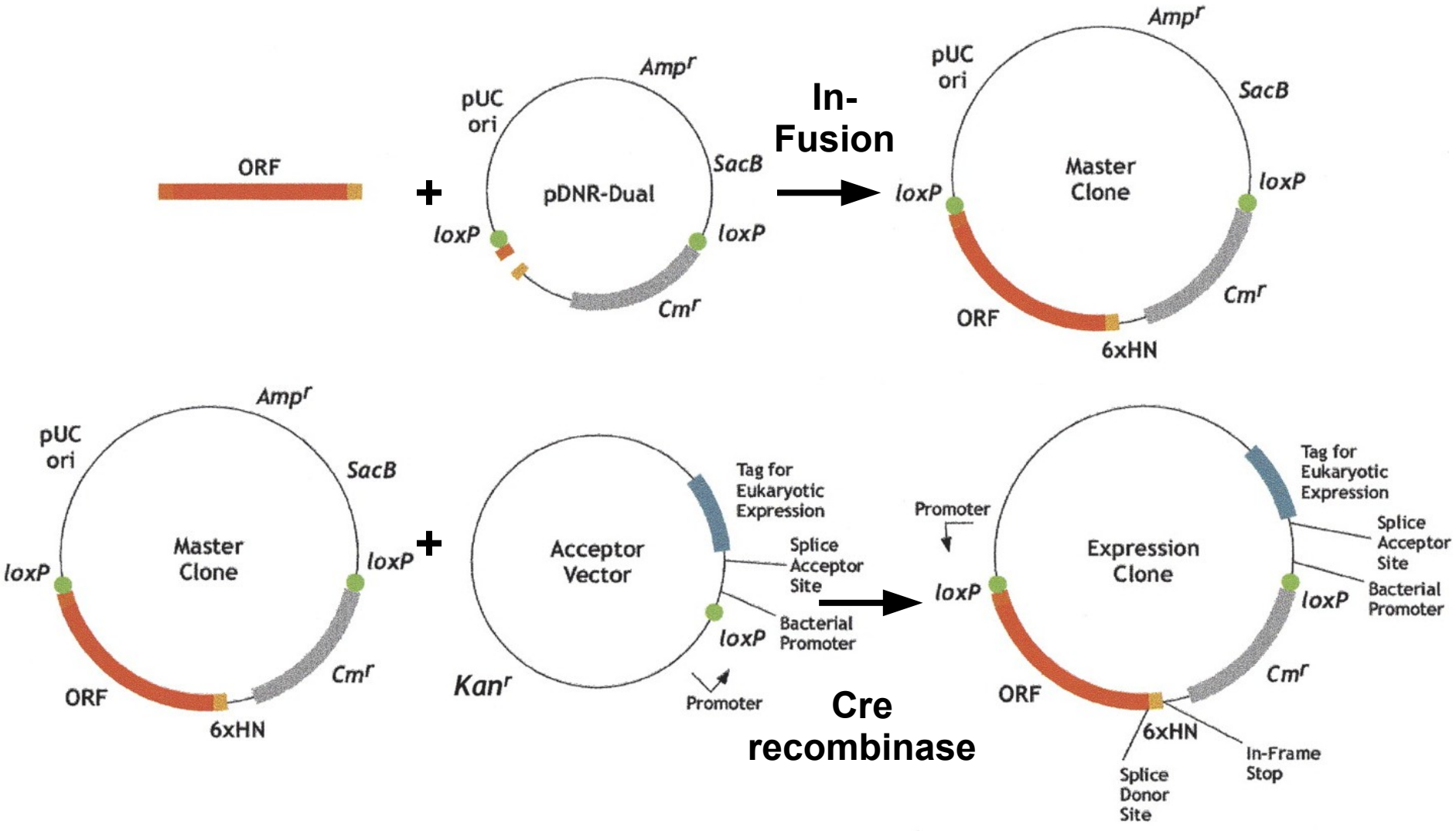


Figure from: Marsischky G & LaBaer J Genome Res. 2004; 14: 2020-2028

Univector site-specific recombination cloning system

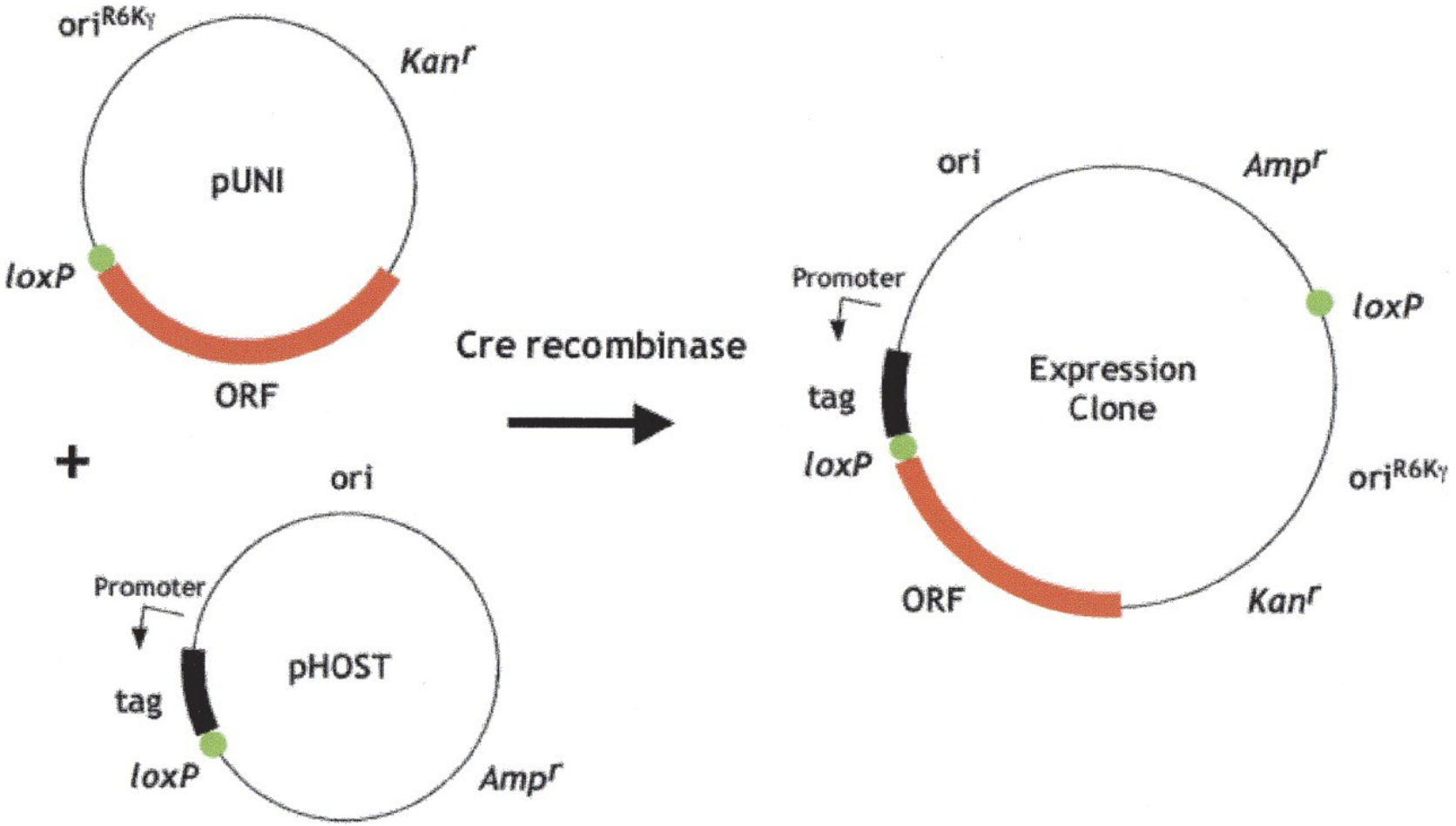
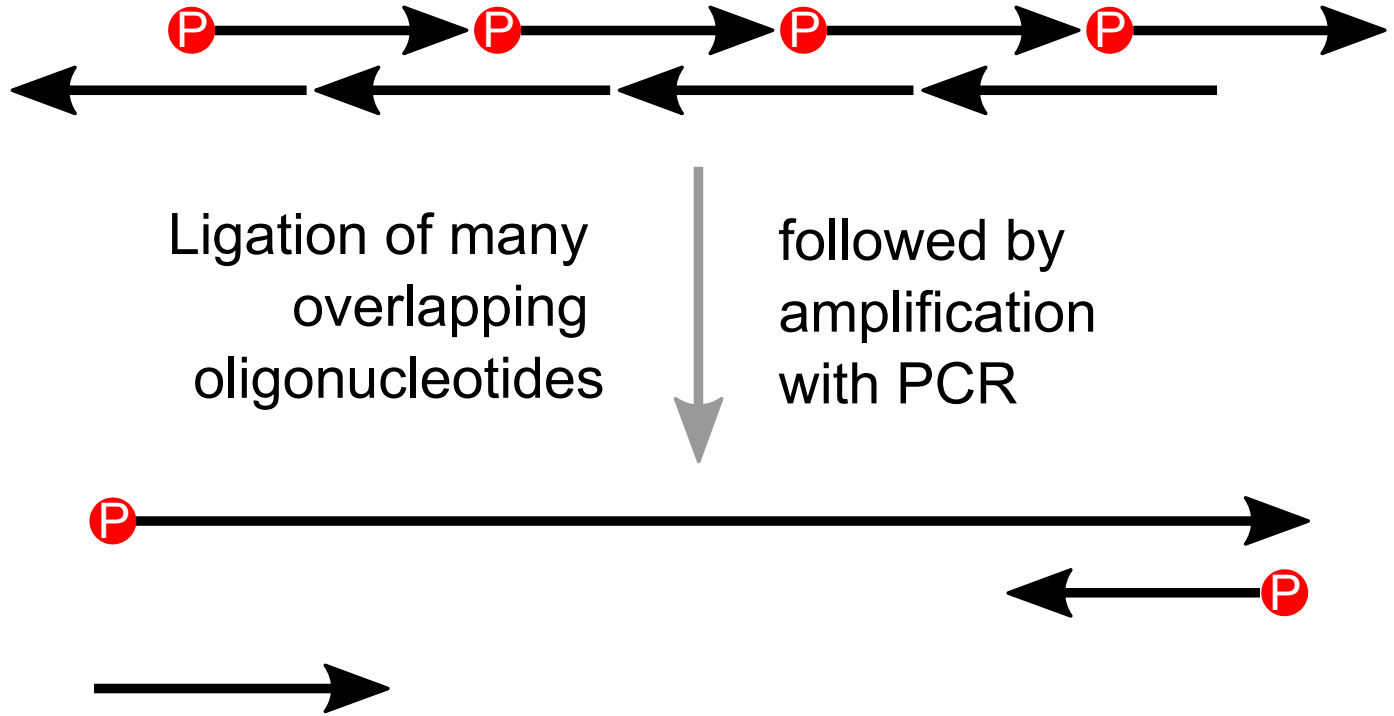


Figure from:
Marsischky G & LaBaer J Genome Res. 2004; 14: 2020-2028

Artificial/De-novo gene synthesis

- No template needed
- Chemical synthesis, max. length 200 bp
- Further assembly by different methods, e.g.

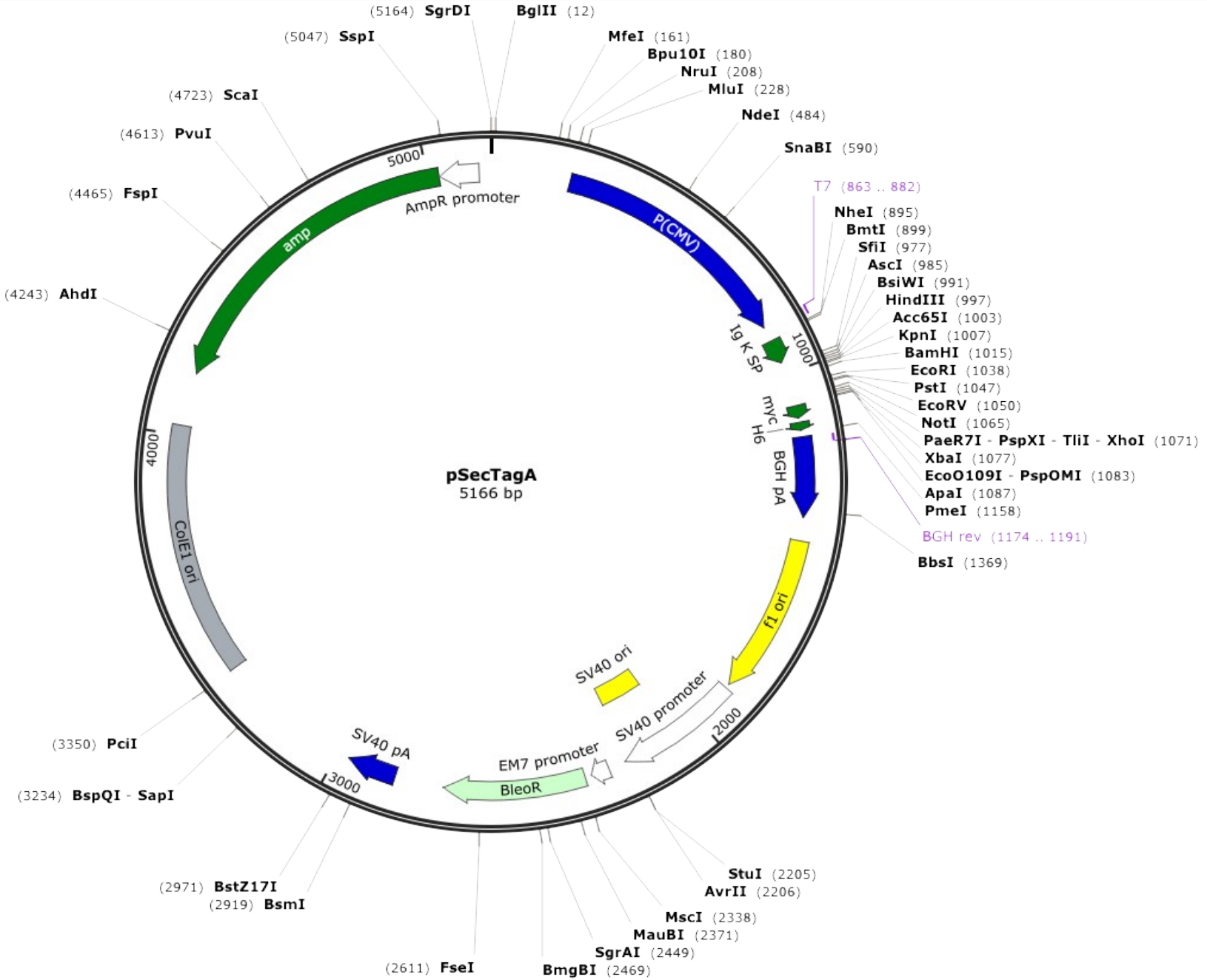


- Sequence verification!
- atm ~0.22€/bp (LifeTechnologies/GeneArt)

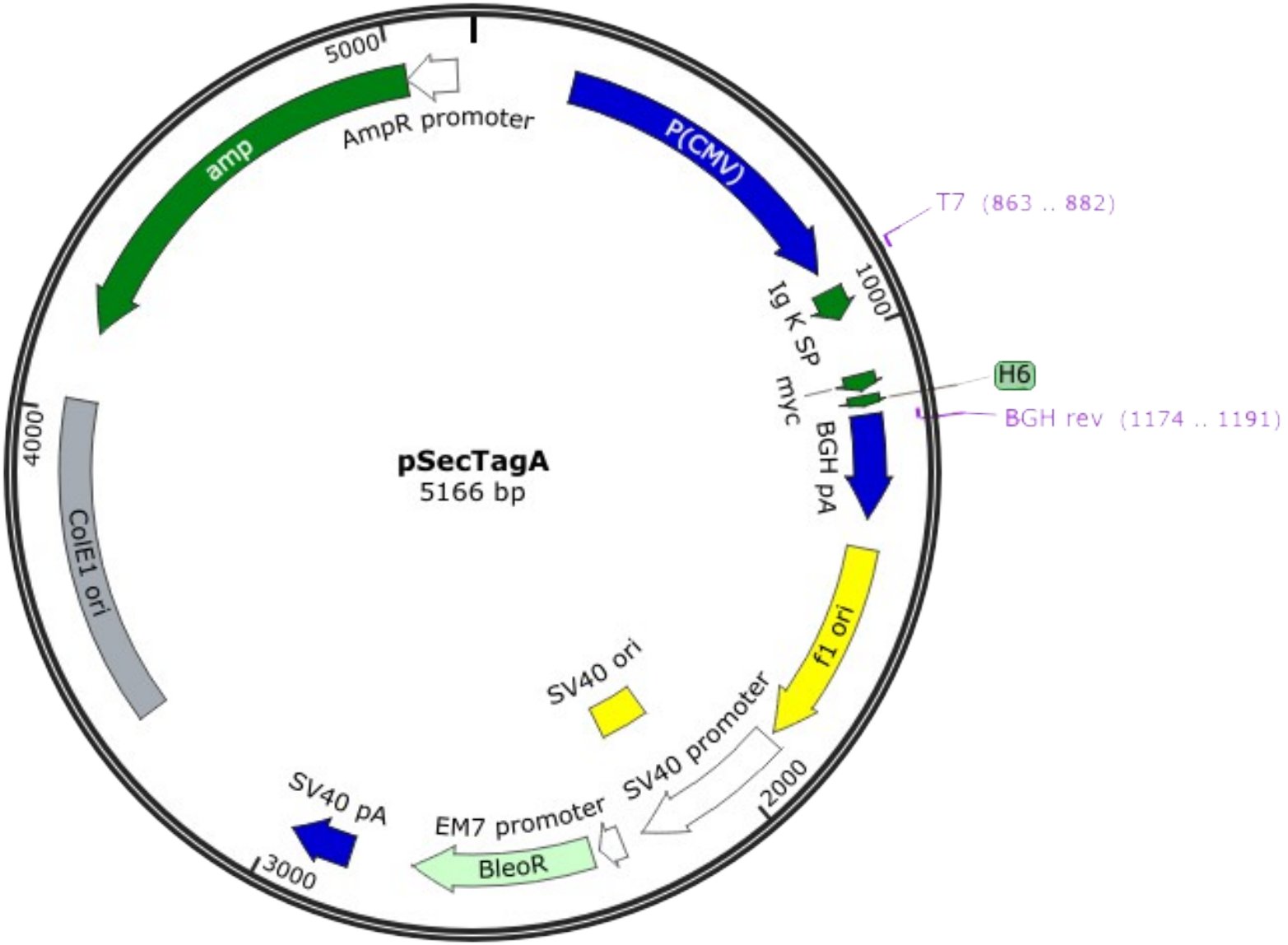
RT-PCR/PCR

- PCR: see lecture 1, slides 18 & 19
- RT-PCR: full-lenght cDNA as a goal; really needed? About 900 intronless genes in the human genome (notably histones and G-protein coupled receptors)
- Non-CDS introns in 3'- and 5'-UTR
- Genes with few exons (e.g. Prox-1: 5 exons (4 CDS, 575 of 737 aa encoded by exon 2)
- Average: 8-9 exons per gene
- Mega-Genes: Titin 35991 aa (363 exons)
- Guide to the human genome: http://www.cshlp.org/ghg5_all/section/gene.shtml

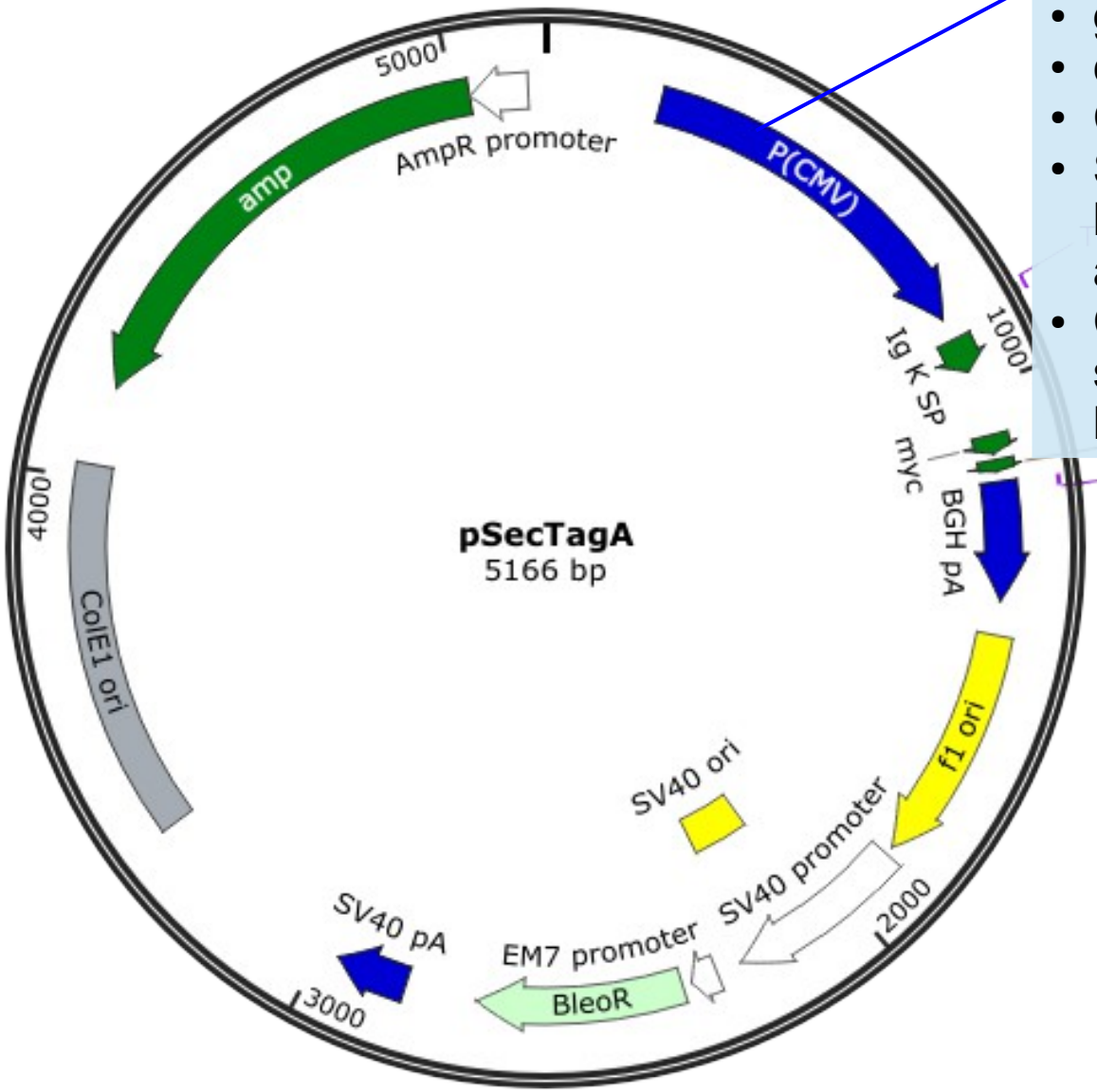
The Anatomy of an Expression Vector



The Anatomy of an Expression Vector



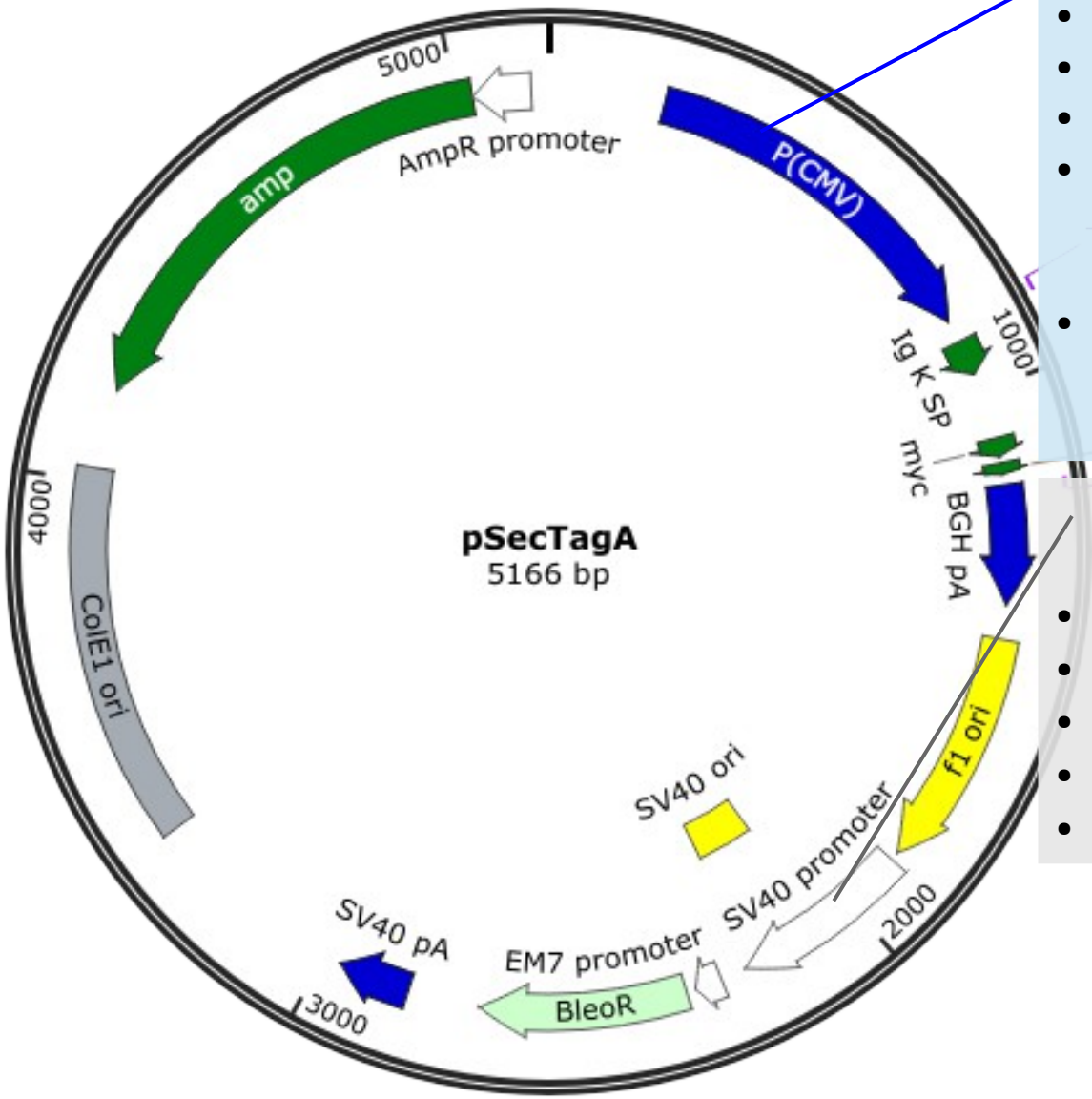
The Anatomy of an Expression Vector



Promoter & polyA for the GOI

- general (mostly viral) promoters
- cell line specific promoters
- Choice important for expression levels
- Sometimes you want low expression levels (better alternative: titrating the amount of DNA used for transfection)
- Constitutive vs. inducible by heat, steroid hormones, heavy metals, hypoxia, interferons, etc.

The Anatomy of an Expression Vector



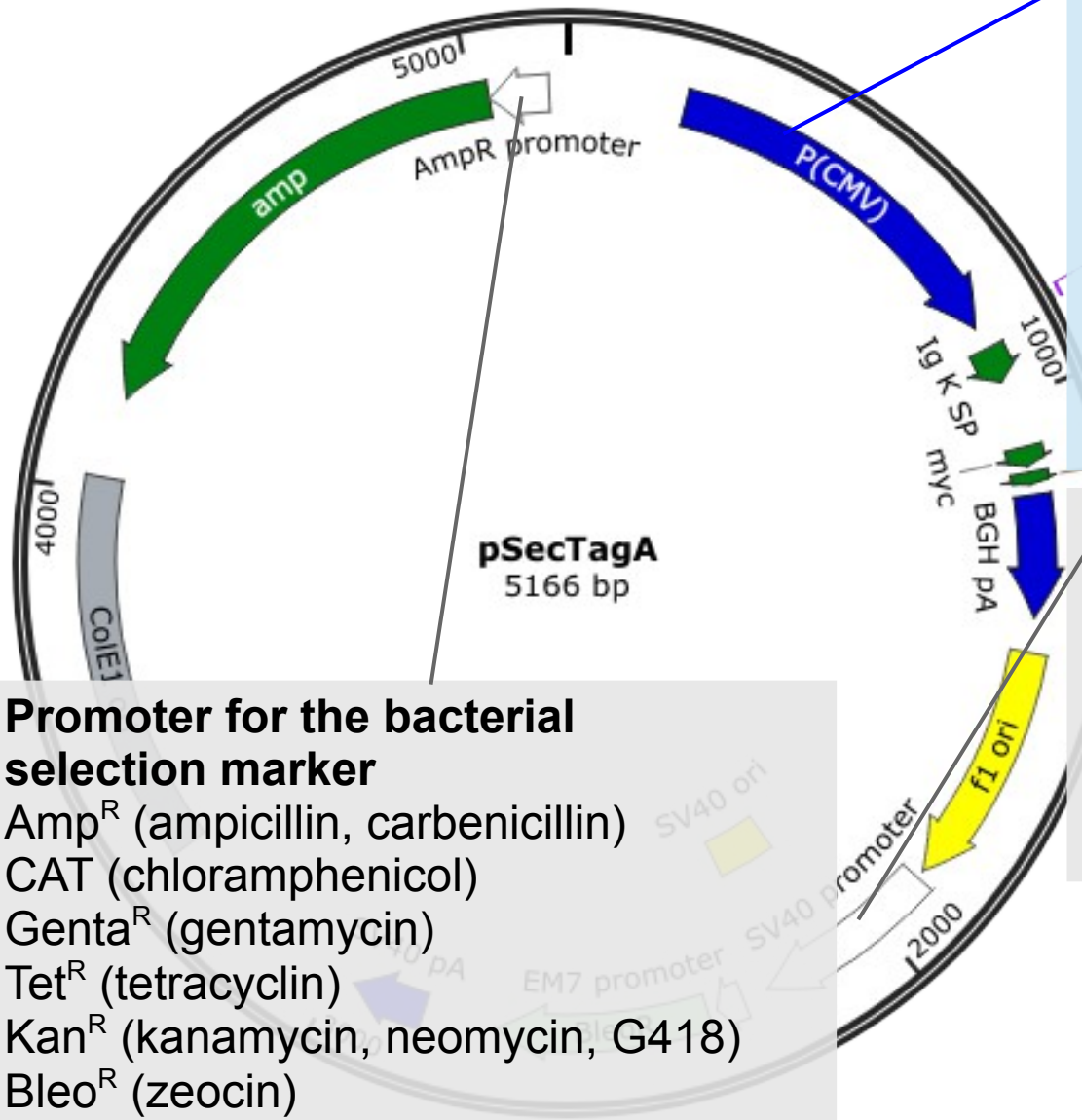
Promoter & polyA for the GOI

- general (mostly viral) promoters
- cell line specific promoters
- Choice important for expression levels
- Sometimes you want low expression levels (better alternative: titrating the amount of DNA used for transfection)
- Constitutive vs. inducible by heat, steroid hormones, heavy metals, hypoxia, interferons, etc.

Promoter & polyA for the mammalian selection marker

- Neo^R (G418)
- $Bleo^R$ (zeocin)
- $Puro^R$ (puromycin)
- $Hygro^R$ (hygromycin)
- Bla^R (blasticidin)

The Anatomy of an Expression Vector



Promoter & polyA for the GOI
- general (mostly viral) promoters
- cell line specific promoters
- Choice important for expression levels
- Sometimes you want low expression levels (better alternative: titrating the amount of DNA used for transfection)
- Constitutive vs. inducible by heat, steroid hormones, heavy metals, hypoxia, interferons, etc.

Promoter & polyA for the mammalian selection marker
- Neo^R (G418)
- $Bleo^R$ (zeocin)
- $Puro^R$ (puromycin)
- $Hygro^R$ (hygromycin)
- Bla^R (blasticidin)

Promoter for the bacterial selection marker
- Amp^R (ampicillin, carbenicillin)
- CAT (chloramphenicol)
- $Genta^R$ (gentamycin)
- Tet^R (tetracyclin)
- Kan^R (kanamycin, neomycin, G418)
- $Bleo^R$ (zeocin)
- Bla^R (blasticidin)

Aminoglycoside antibiotics

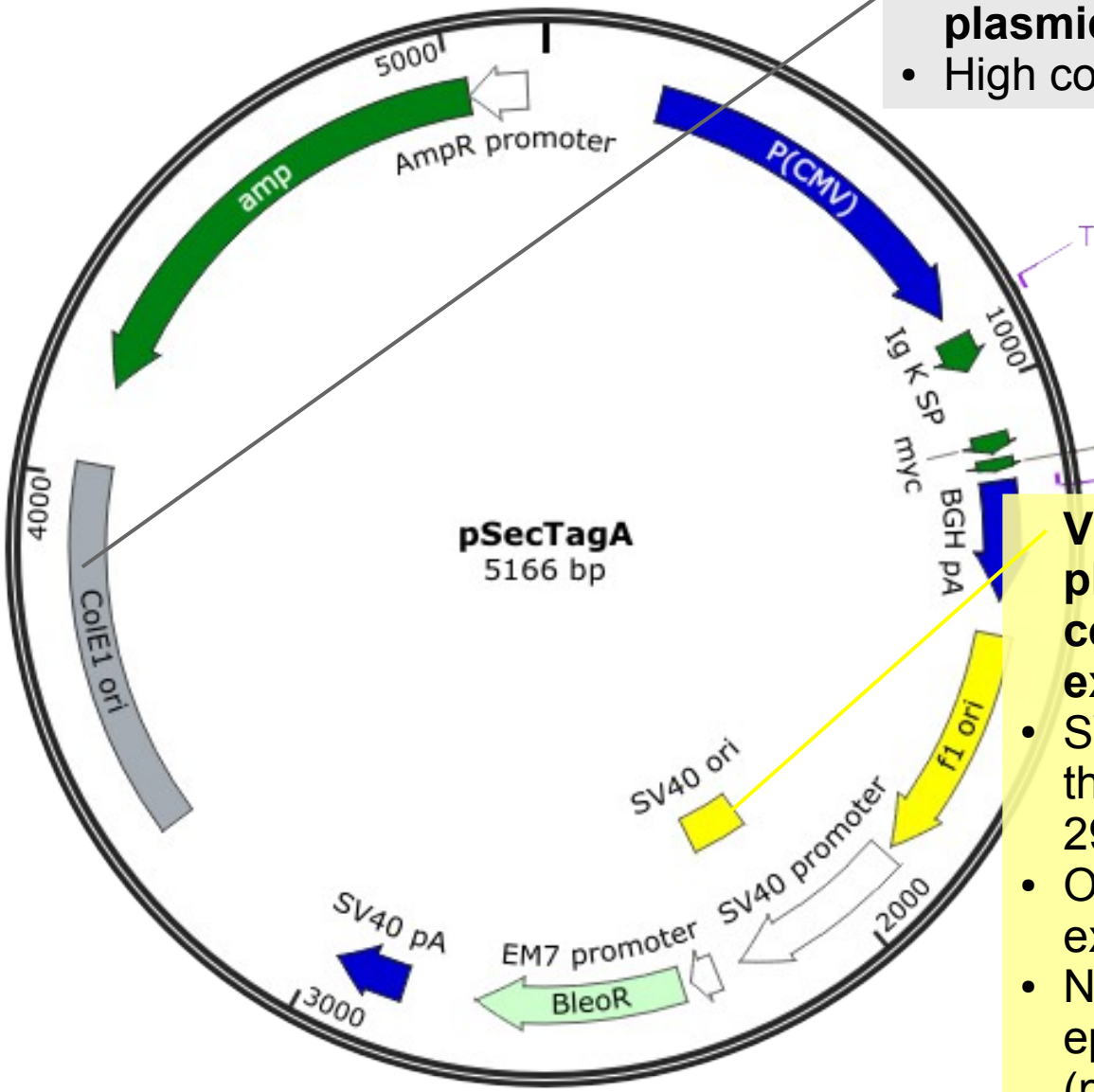
antibiotic	Primary target	typical effective concentration for bacteria	typical effective concentration for mammalian cells	~Price (€/25g)	used as a human drug
G418 (geneticin)	Prokaryotic & eukaryotic ribosome	5 µg/ml	100-1500 µg/ml	2330	no
neomycin	Prokaryotic ribosome	150 µg/ml	-	45	yes
kanamycin	Prokaryotic ribosome	25 µg/ml	-	412	yes
paromomycin	Prokaryotic & eukaryotic ribosome	?	?	?	yes

- KanR and NeoR are the same thing!
- There are several different KanR/NeoR resistance genes (from different origins), but they all code for aminoglycoside phosphotransferases and all inactivate G418/ neomycin/kanamycin/paromomycin
- The difference is the promoter (mammalian promoter → neo^R, bacterial promoter → $kana^R$)
- One kana/neo or zeo resitance gene can be used with dual (bacterial/mammalian) promoters to confer resistance in both E.coli and mammalian cells (smaller plasmid size for “shuttle vectors”)
- No G418 in 293T cells (293T cells have been stably transfected with a large T antigen-expressing vector, which carries neoR)

The Anatomy of an Expression Vector

Bacterial origin of replication allows plasmid replication in E.coli

- High copy or low copy



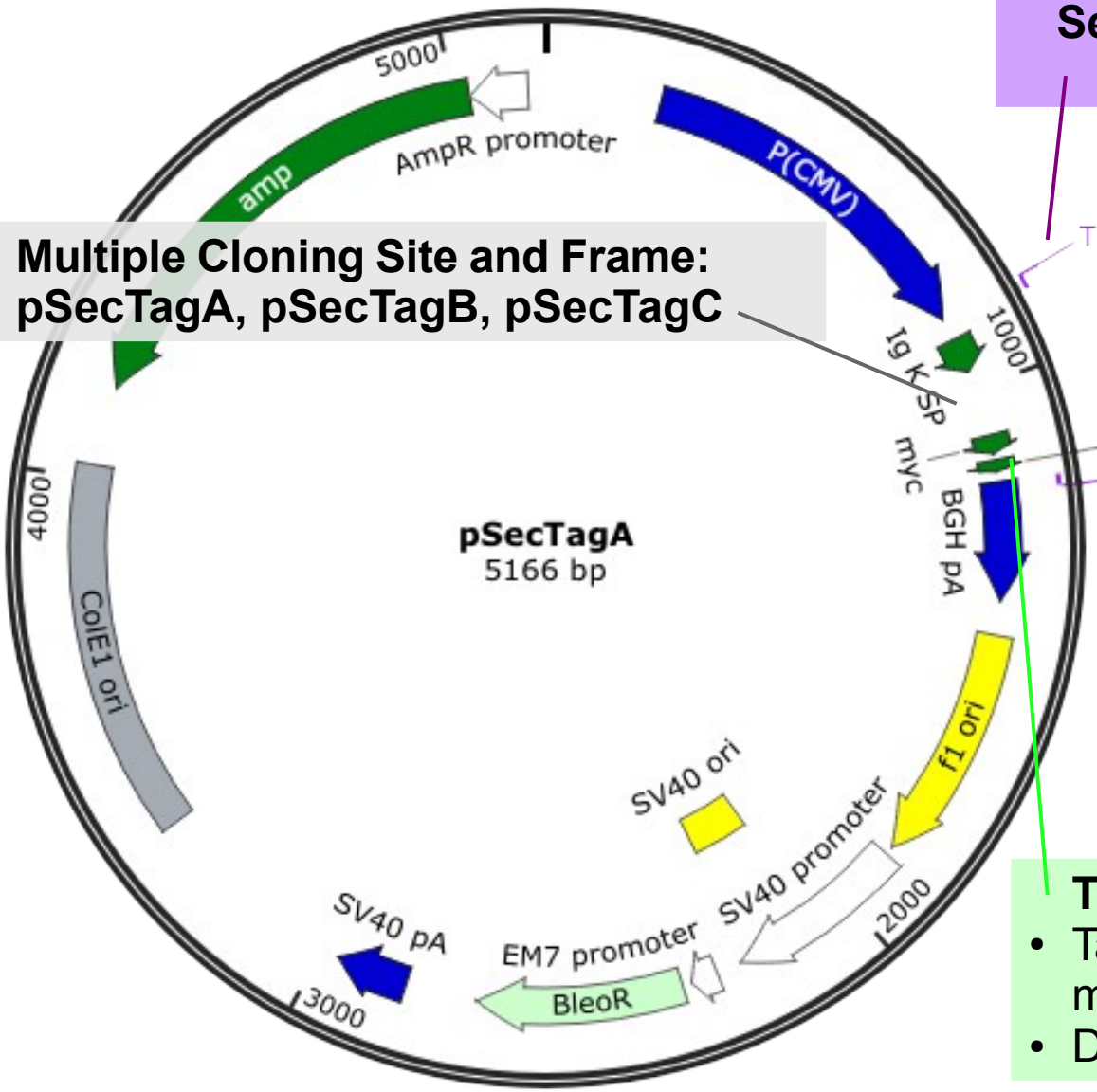
Viral origin of replication enables plasmid to replicate in mammalian cells resulting in very high expression levels

- SV40 ori (works with cells that express the SV40 large T antigen: Cos-7, 293T)
- OriP (works with human cells that express EBNA-1: 293EBNA)
- No stable cell lines possible for these episomally maintained plasmids (primary cells can contain viruses!)

The Anatomy of an Expression Vector

Sequencing primer annealing sites

Multiple Cloning Site and Frame:
pSecTagA, pSecTagB, pSecTagC



T7 (863 .. 882)

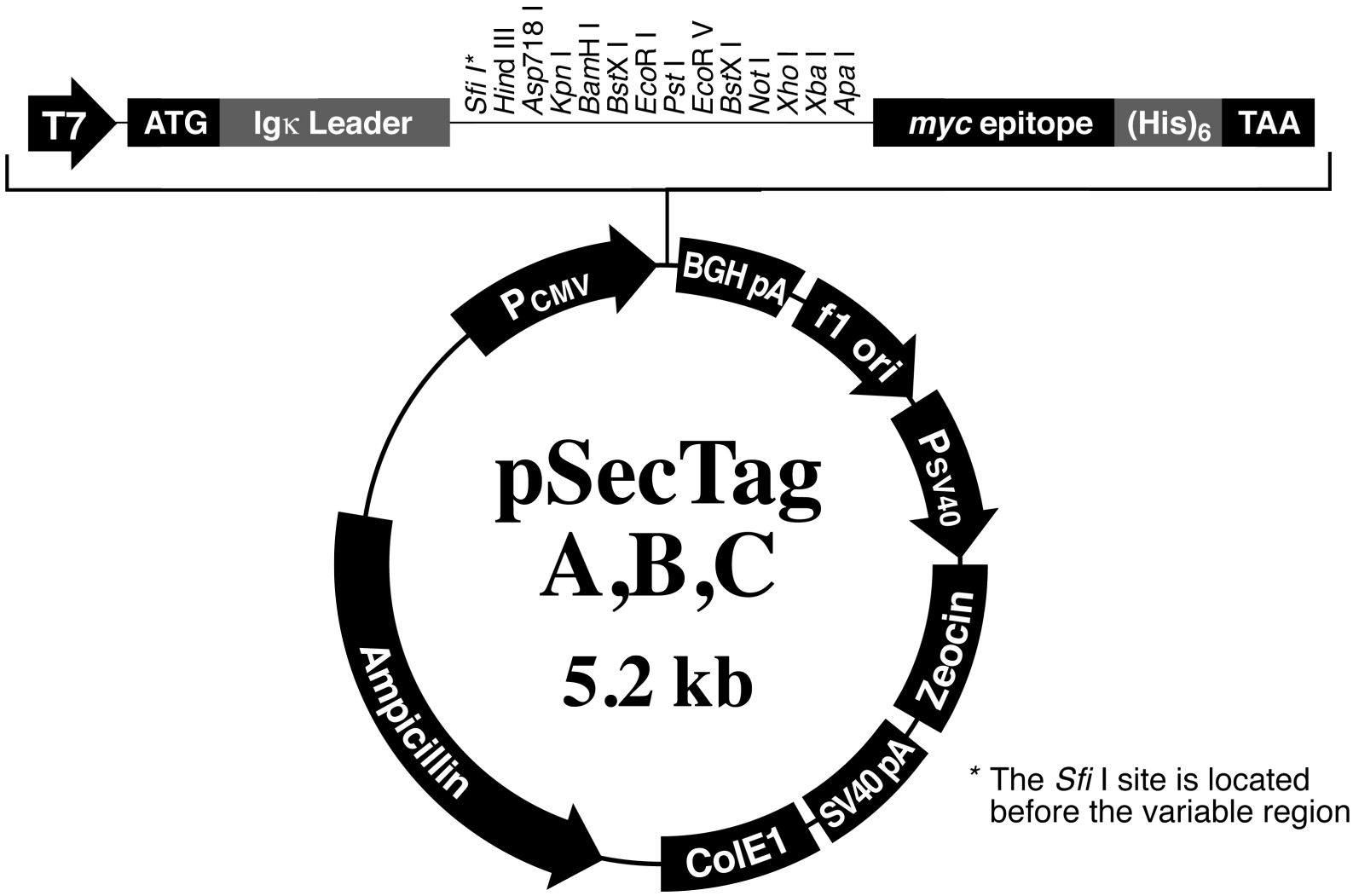
H6

BGH rev (1174 .. 1191)

Tags and targeting signals

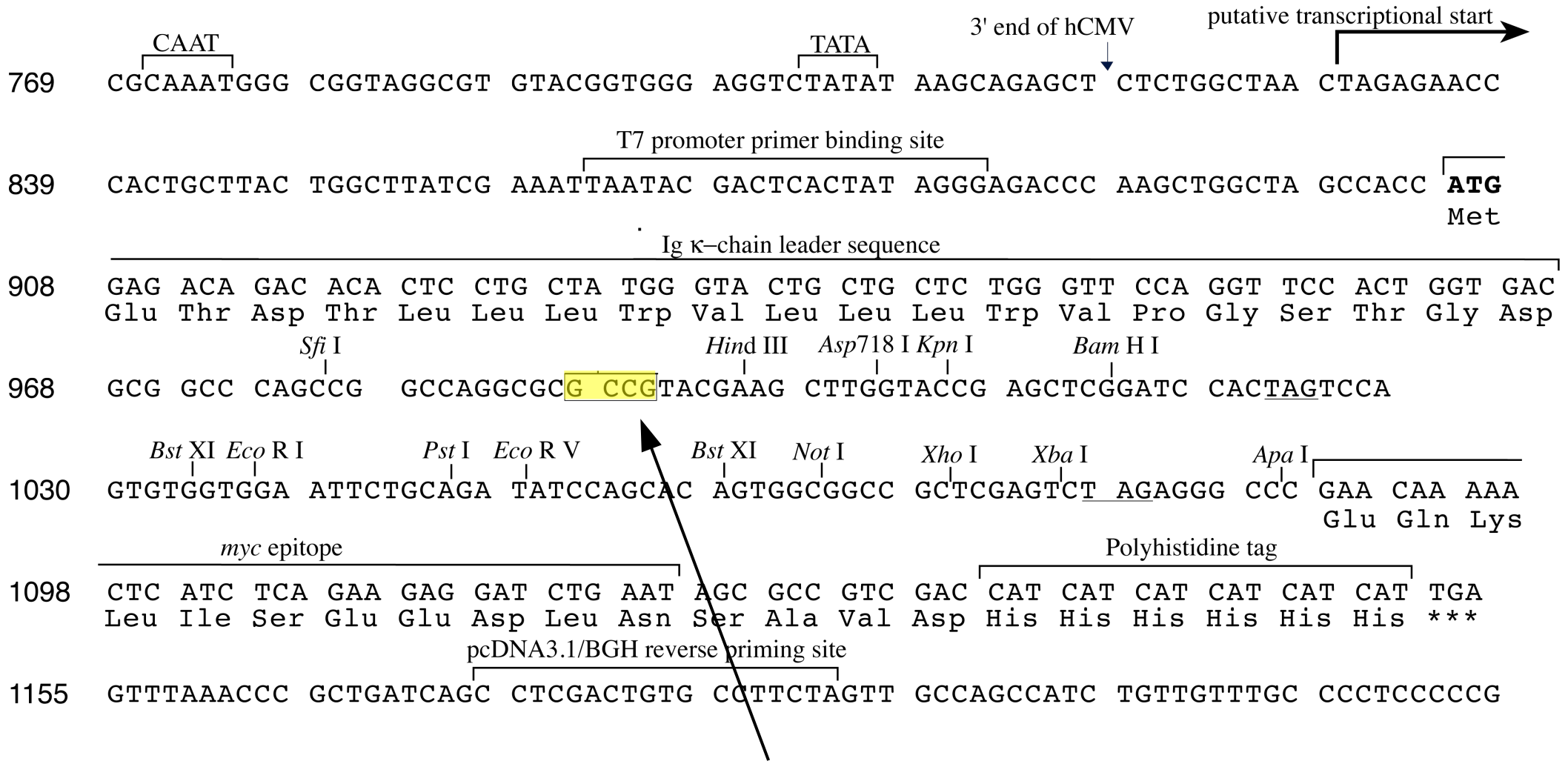
- Targeting signals: secretion, membrane anchoring, etc.
- Detection and purification tags

Multiple Cloning Site & Frame

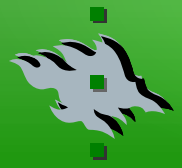


* The *Sfi* I site is located before the variable region

Multiple Cloning Site & Frame



Different number of nucleotides in the yellow area for A/B/C version of the vector to enable in-frame cloning for each possible reading frame



Literature

- **Many Paths to Many Clones: A Comparative Look at High-Throughput Cloning Methods**
  Gerald Marsischky and Joshua LaBaer *Genome Res.* 2004 14: 2020-2028
  http://genome.cshlp.org/content/14/10b/2020.long

- **Mammalian Expression Vectors**
  LifeTechnologies
  http://tools.lifetechnologies.com/content/sfs/productnotes/F_071215_MammalianExpressionVectors.pdf