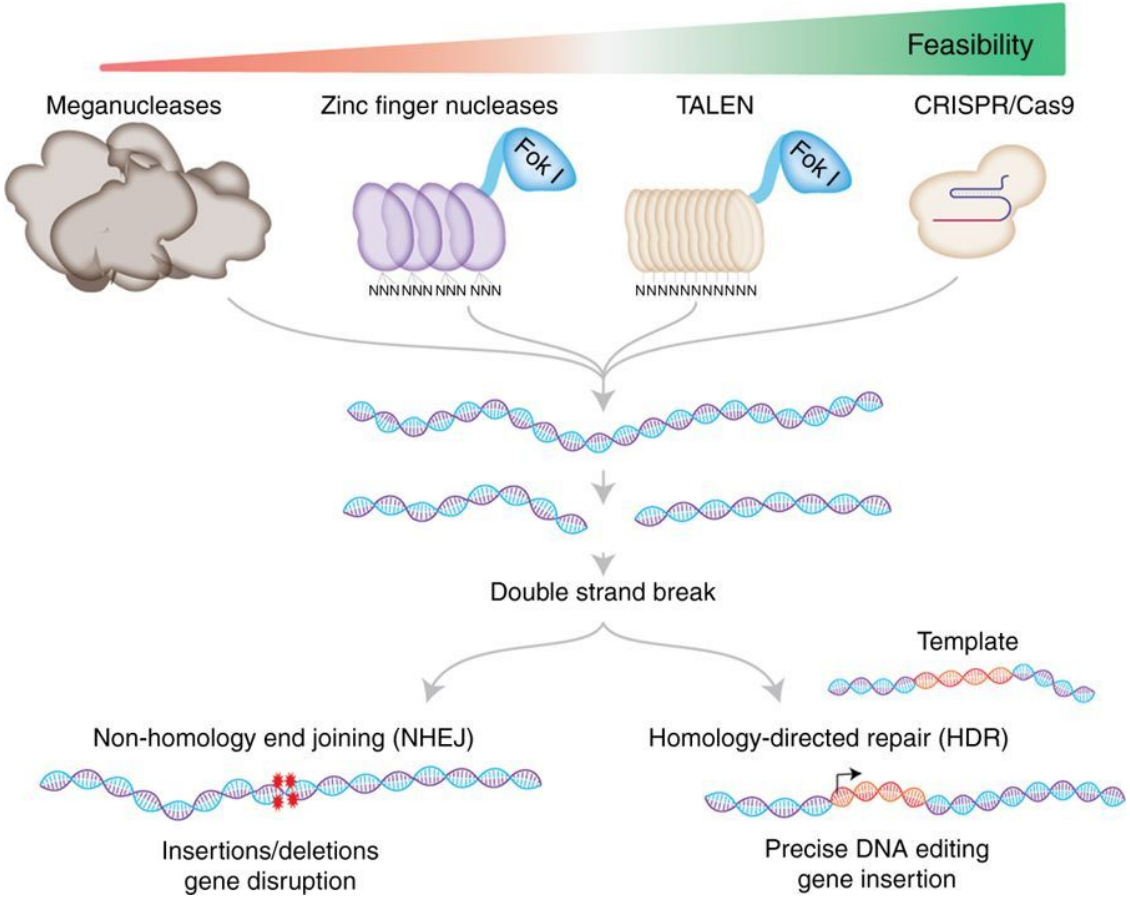
Lecture 28: Genome editing

Overview



Meganucleases (engineered restriction enzymes) recognize long stretches of DNA sequences.

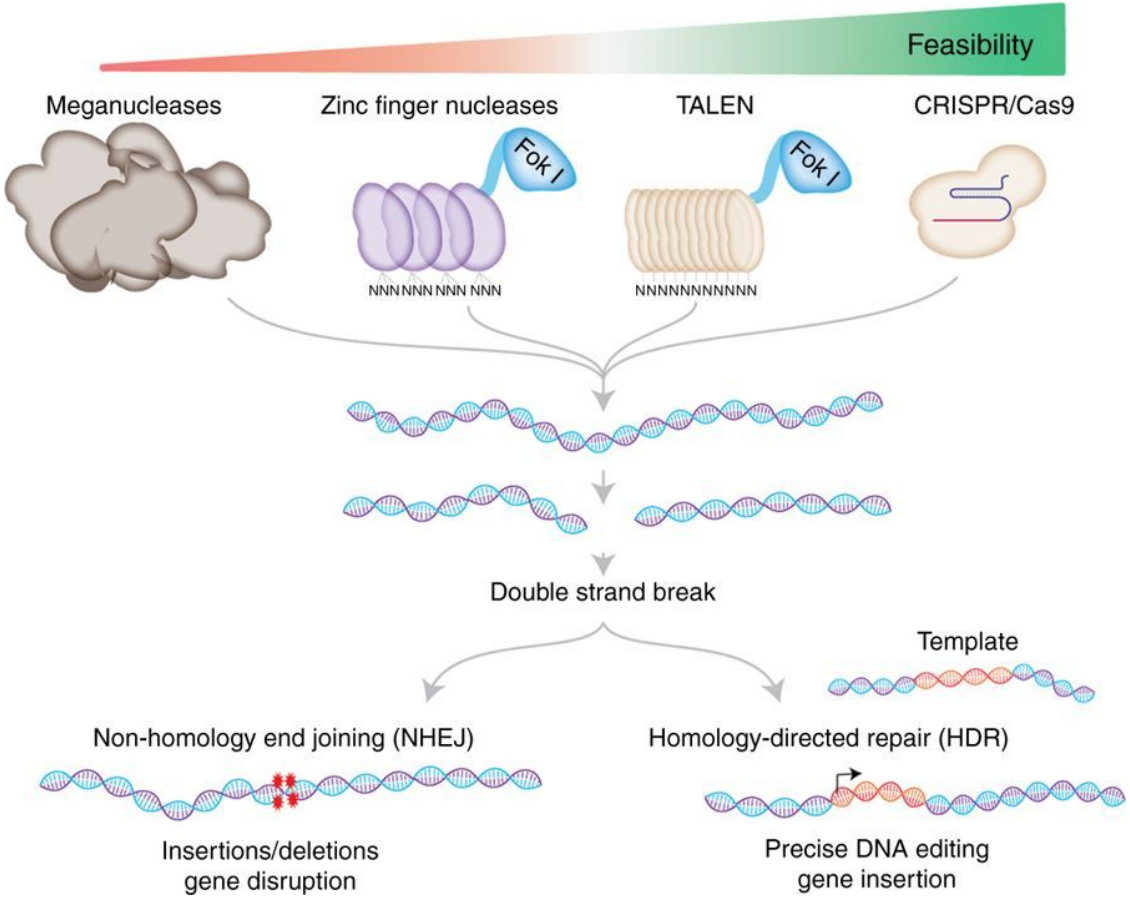
Each zinc finger nuclease recognizes triple DNA code.

Each TALE (transcription activator-like effector) recognizes an individual base.

CRISPR (clustered regularly interspaced short palindromic repeat) recognizes target by simple RNA–DNA base pairing and the PAM (protospacer-adjacent motifs) sequence.

Adli (2014) Nat. Comm.

Overview



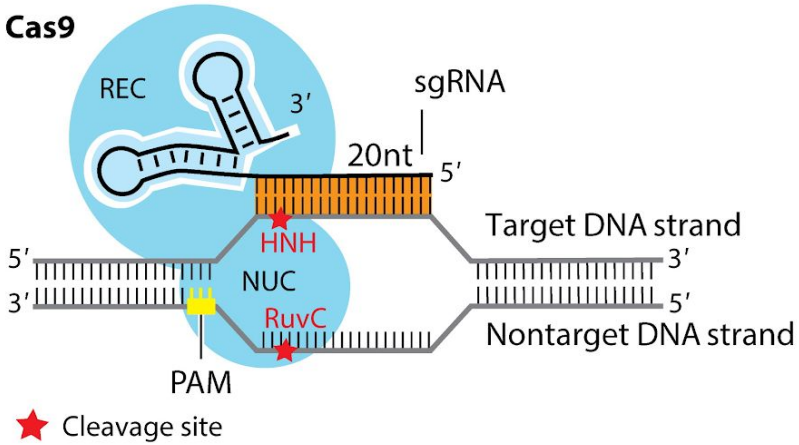
All result in DNA double-strand breaks, which are repaired either by error-prone non-homology end joining (NHEJ) or homology-directed repair (HDR).

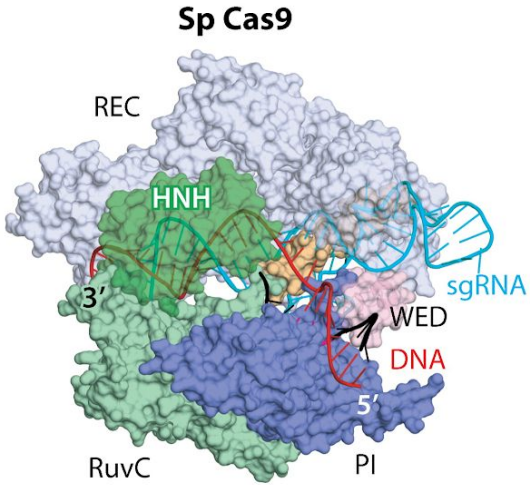
NHEJ results in random indels and gene disruption at the target site.

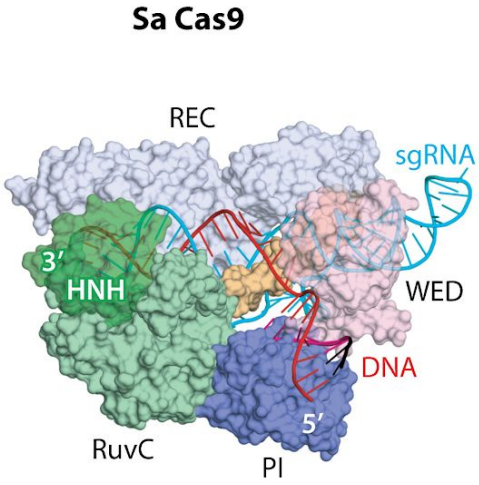
HDR can be harnessed to insert a specific DNA template (single stranded or double stranded) at the target site for precise gene editing.

Adli (2014) Nat. Comm.

Guide-RNA design





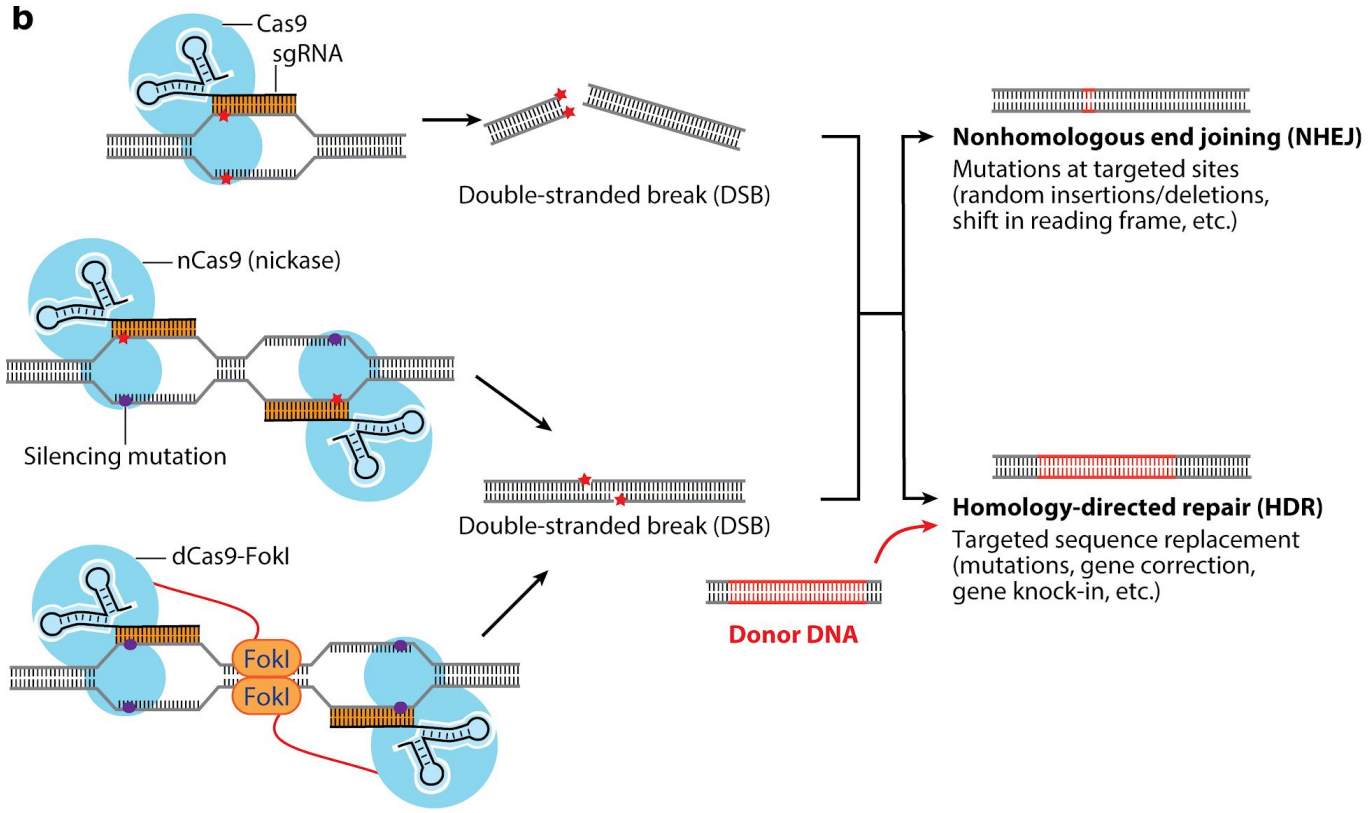


Cas9 is a naturally evolved, RNA-guided nuclease.

- Recognizes its target DNA through ~20 nt base-pairing interaction b/w a single guide RNA (sgRNA).
- Interacts with the protospacer-adjacent motif (PAM) of its DNA target through its PAM-interacting (PI) domain at its C terminus.
- Uses its two nuclease domains (HNH and RuvC) to cleave the double-stranded DNA, creating a DSB.

Wang (2016) Ann. Rev. Biochem.

Guide-RNA design

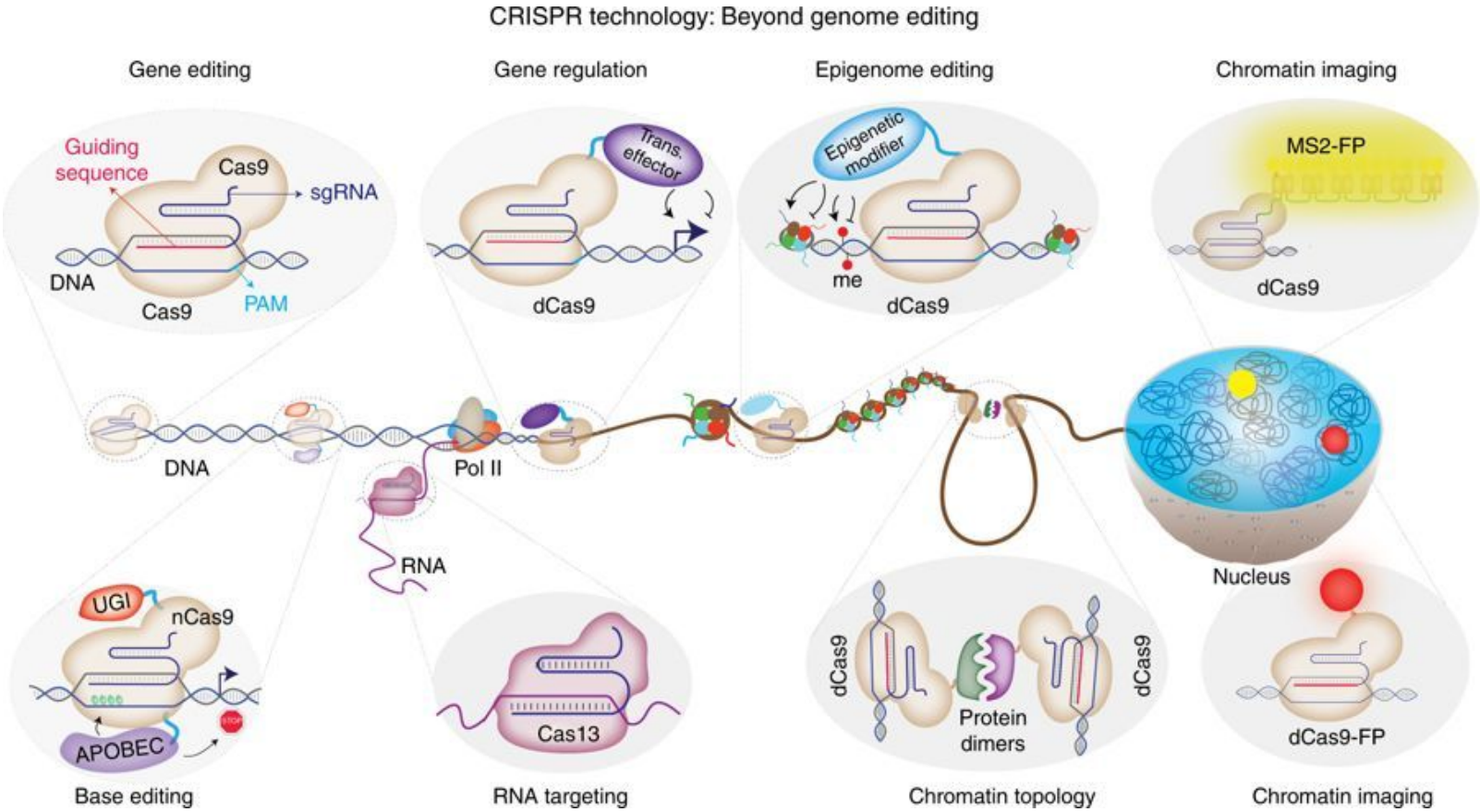


A mutation in one nuclease domain of Cas9 creates a Cas9-based nickase (nCas9) that cleaves only one strand of DNA.

Can increase specificity by using a pair of nCas9s that target each strand of DNA at adjacent sites.
Both nCas9–sgRNA complexes must be present at the target site for DSB creation.

Wang (2016) Ann. Rev. Biochem

Overview

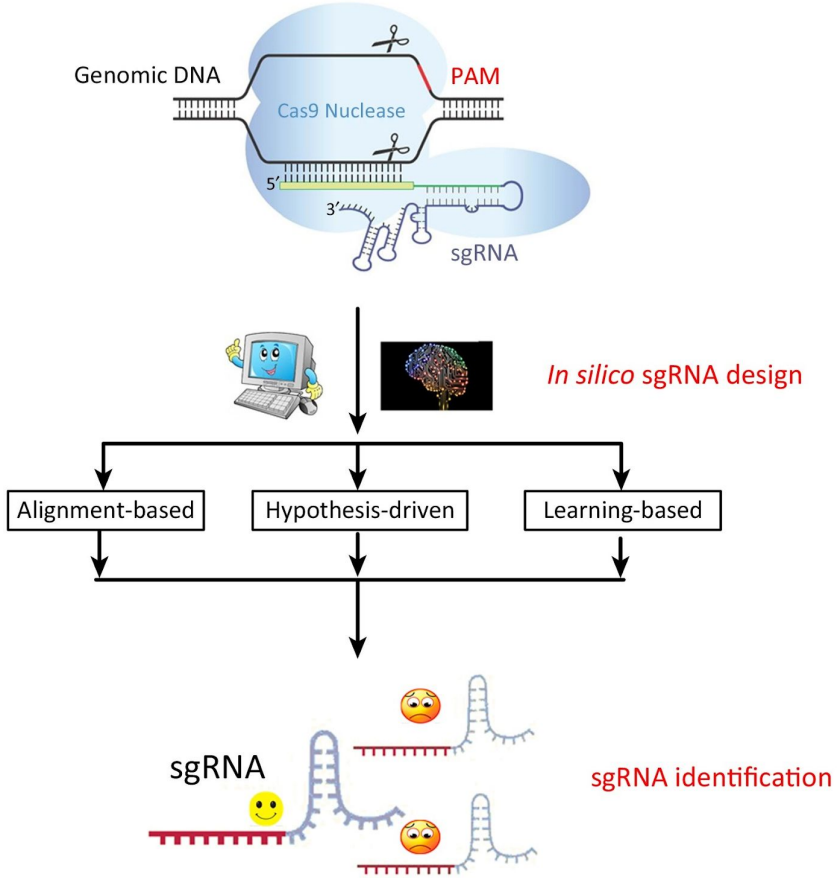


Adli (2014) Nat. Comm.

Guide-RNA design

All of these approaches require the accurate and efficient targeting of the CRISPR-Cas9 system to the desired location.

- Correct identification of the optimal target-site and subsequent design of the complimentary gRNA.
- Generic target libraries may not be well-suited for specific research purposes. The design of custom gRNAs is hence frequently required.
- Maximize on-target activity (guide efficiency) while also minimizing potential off-target effects (guide specificity).

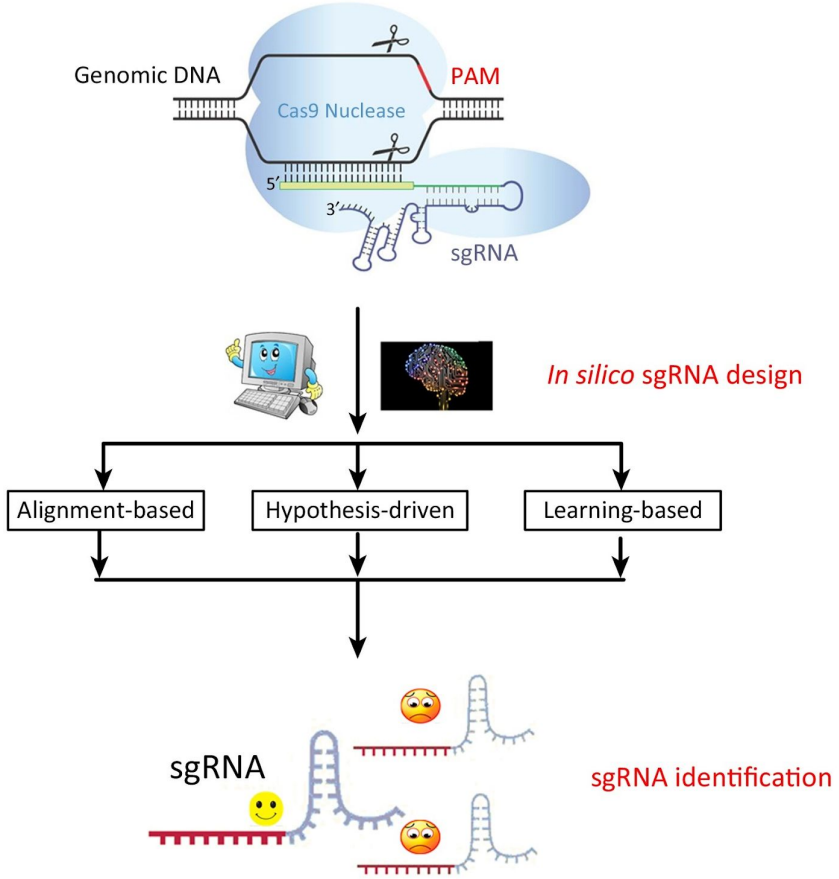


Chuai (2016) Trends. Biotech.; Wilson (2018) Front. Pharmocol.

Guide-RNA design

Important features for predicting **on-target activity**:

- position-specific nucleotides, such as a G preceding the PAM being a strong indicator of CRISPR-Cas9 activity; region adjacent to the PAM, a region that has become known as the seed region.
- global variables: GC content, gRNA melting temperature.
- position of the target site relative to the TSS and position within the protein.



Wilson (2018) Front. Pharmocol.

Guide-RNA design

Program name	Identifies/Scores targets?	Training Data used for model	gRNA transcription method	Model Features and Implementation	Comments
SSFinder (Chari et al., 2015)	Identifies targets	NA	NA	NA	Does not score activity. Filters targets if the seed region (12bp preceding the PAM) are not unique within the input sequence.
sgRNAcas9 (Xie et al., 2014)	Identifies targets	NA	NA	NA	Does not score activity. Ranks targets based on potential off-targets (detected via alignment with SeqMap)
CRISPRseek (Zhu et al., 2014)	Identifies targets	NA	NA	NA	Does not score activity. Implemented as a Bioconductor package for R.
sgRNA Scorer (Chari et al., 2015)	Identifies and scores targets	**Mutation rates at target sites** in HEK293t cells following treatment with CRISPR-Cas9 and gRNAs	*In vitro* transcription using a U6 promoter	SVM model using nucleotide composition of target site of the target site	Standalone program.
Rule Set 1 (Doench et al., 2014)	Scores targets	**Enrichment rates of transfected gRNA** following selection for changes in expression of cell-surface markers as determined by FACS in human and mouse cells.	*In vitro* transcription using a U6 promoter	Logistic Regression model using nucleotide composition of the target site	No implementation is available however the feature weights are provided in the paper allowing for standalone implementation.
CRISPRscan (Moreno-Mateos et al., 2015)	Identifies and scores targets	Mutation rates at target sites in zebrafish embryos.	*In vitro* transcription using a T7 promoter	Linear Regression model using nucleotide composition of the target site	Implemented as a web-app (www.crisprscan.org). Features weights are also provided in the paper allowing for standalone implementations.
WU-CRISPR (Wong et al., 2015)	Identifies and scores targets	Reanalysis of the data from Doench 2014	*In vitro* transcription using a U6 promoter	SVM model using nucleotide composition of the target site and sgRNA secondary structure	Employs strict filtering criteria which results in a majority of potential targets being discarded prior to scoring. Implemented as a standalone program.
Azimuth (Doench et al., 2016)	Identifies and scores targets	Enrichment rates of transfected sgRNAs targeting drug-resistance pathways following drug challenge in human cells. Also incorporates data from Doench 2014	*In vitro* transcription using a U6 promoter	Combined SVM and Logistic Regression model using nucleotide composition of the target site and flanking region, sgRNA secondary structure and position of target site relative to transcription start	Implemented as a stand-alone program and as a Web-app (https://www.microsoft.com/en-us/research/project/crispr/)
TUSCAN (Wilson et al., 2018)	Identifies and scores targets	Reanalysis of the data from Chari 2015	*In vitro* transcription using a U6 promoter	Random Forest model using nucleotide composition of the target site and flanking region	Provides both activity score and general active/inactive classification. Implemented as a stand-alone program and as a Web-app (https://www.gt-scan.net/tuscan)

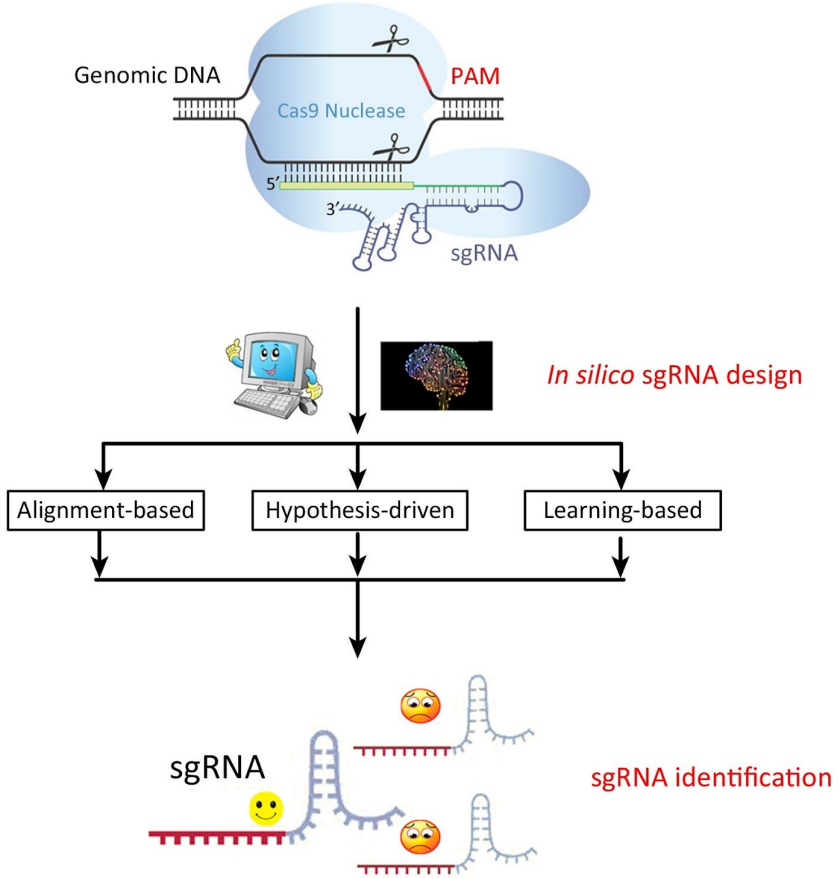
Wilson (2018) Front. Pharmocol.

Guide-RNA design

Predicting **off-target activity**:

- Alignment of the short target sequences is typically achieved using Bowtie and BWA (better suited for handling short sequences compared to BLAST).
- But, they miss high-mismatch off-targets but even some with only one mismatch.
- May be use bi-directional aligners.

Not every putative off-target is actually functional: many FPs.



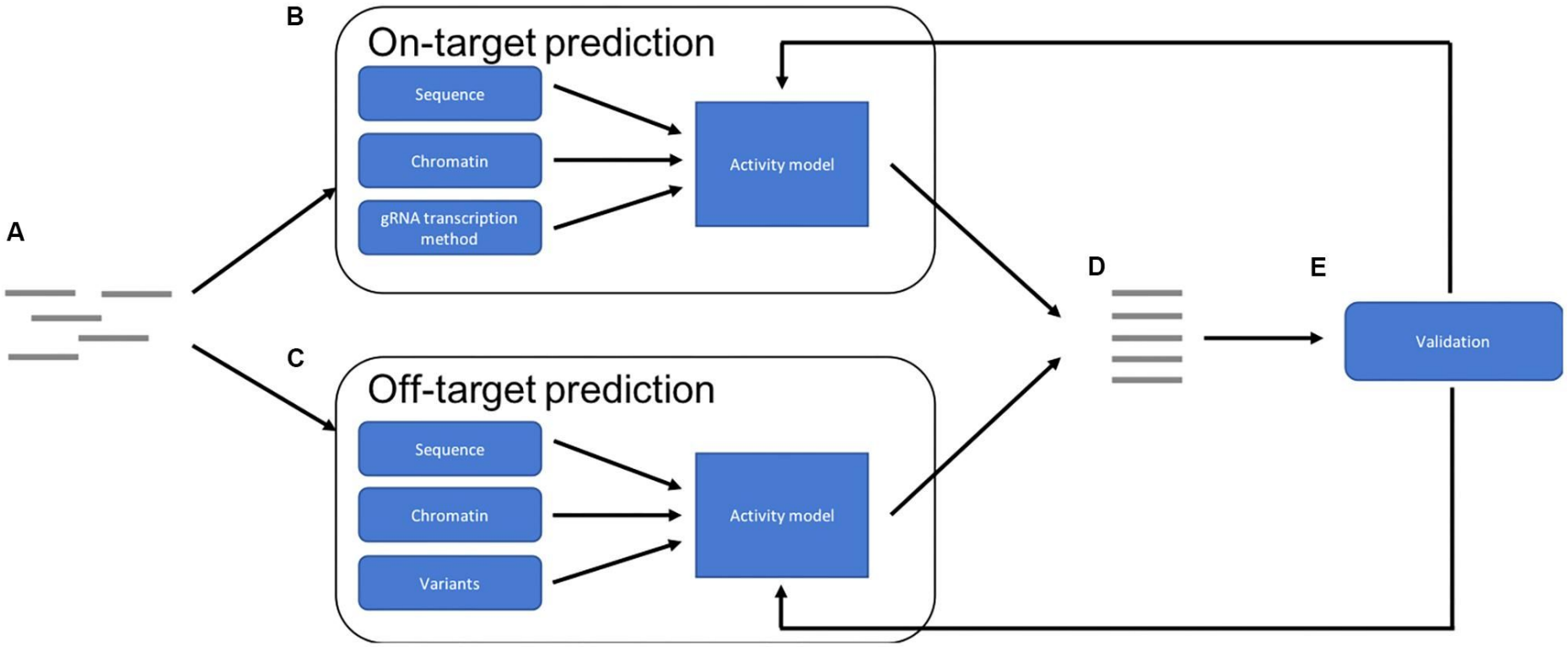
Wilson (2018) Front. Pharmocol.

Guide-RNA design

Program name	Off-target detection method	Off-target scoring method	Notes
GT-Scan (O'Brien and Bailey 2014)	Alignment with Bowtie2	NA	Allows user to identify the required sequence rules (e.g., what PAMs are acceptable, how large a target site). Available as a Web-app (https://www.gt-scan.net/)
CCTop (Stemmer et al. 2015)	Alignment with Bowtie	Custom model based on presence of mismatches and weather they fall within seed region.	Allows for selection of various canonical PAMs and some sequence limitations. Available as a Web-app (https://crispr.cos.uni-heidelberg.de/)
CROP-IT (Singh et al. 2015)	Alignment with PATMAN	Custom model based on position of mismatches and also weather the potential off-target falls within a DNase-sensitive region	Allows for selection between NGG or NNG PAMs. Available as a Web-app (http://cheetah.bioch.virginia.edu/AdliLab/CROP-IT/homepage.html)
CRISPOR (Haeussler et al. 2016)	Alignment with BWA	MIT-Broad score	Allows for selection of various canonical PAMs. Available as a Web-app (http://crispor.tefor.net/)
Elevation (Listgarten et al. 2018)	Custom alignment tool	Custom model based on the number, position and type (wobble vs bulge) of mismatches.	Available as a web-app (crispr.ml). The scoring method is available as a stand-alone tool
CRISTA (Abadi et al. 2017)	NA	Custom model based on the number, position and type (wobble vs bulge) of mismatches.	Available as a stand-alone tool

Wilson (2018) Front. Pharmocol.

Guide-RNA design



Wilson (2018) Front. Pharmocol.