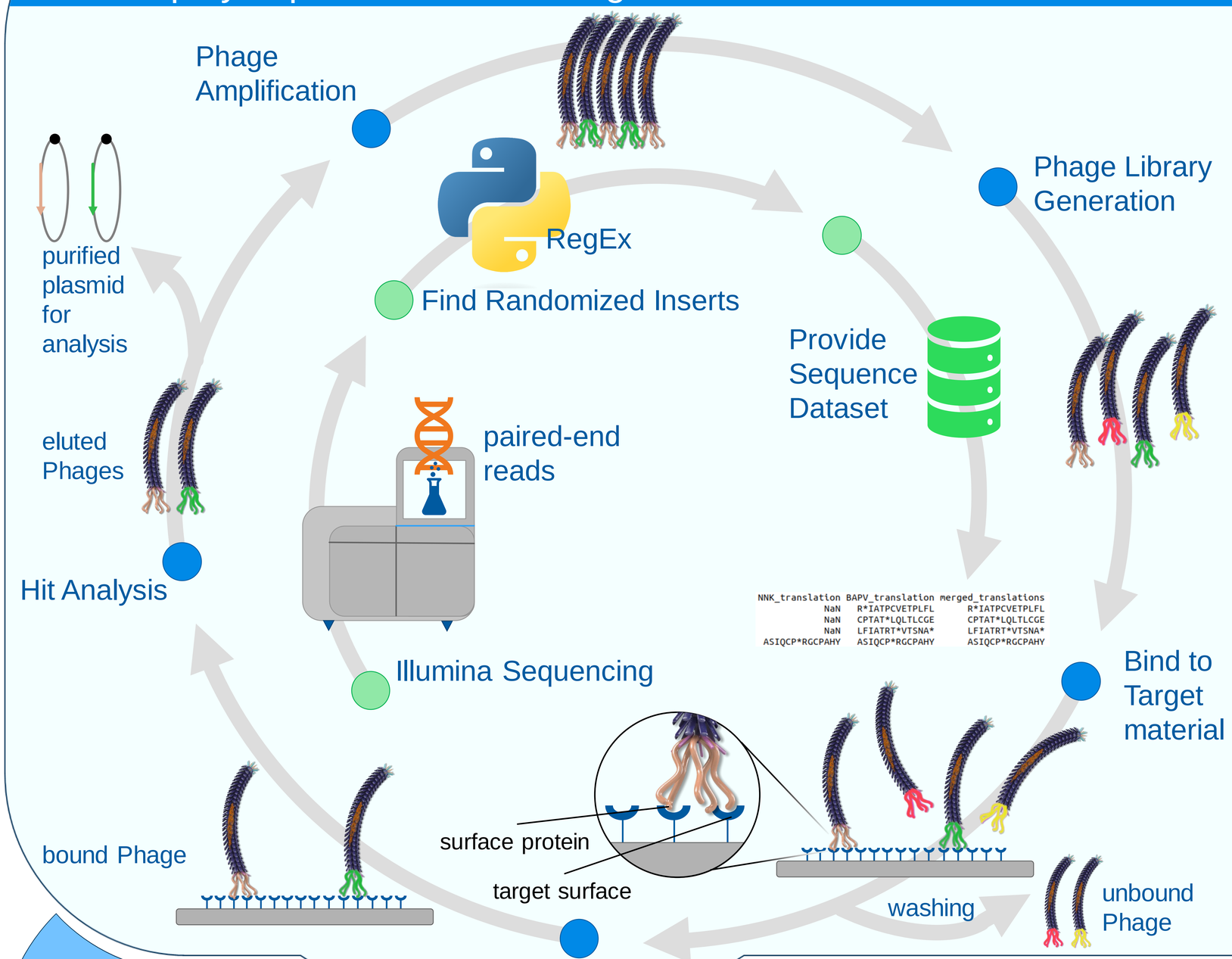
Comparative analysis of next-generation sequencing data generated from phage display trails with inorganic materials

Christoph Bloß[a], Tina Kießlich[a], Christian Hintersatz[a], Franziska Lederer[a]
[a] Helmholtz Institute Freiberg for Resource Technology, Department Biotechnology

HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

HELMHOLTZ INSTITUTE FREIBERG FOR RESOURCE TECHNOLOGY

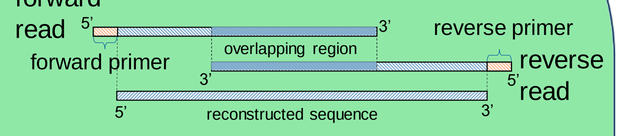
Objective: The aim of this work is to find meaningful and statistical proven ways to identify metal binding peptides in multiple trail phage display experiments with inorganic materials



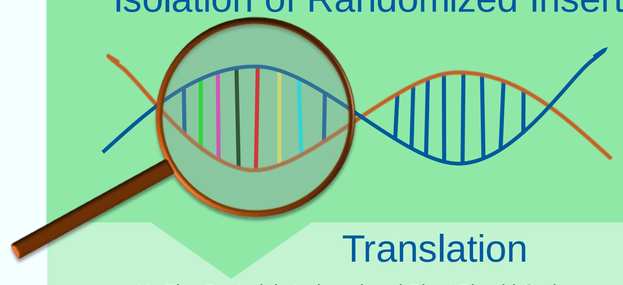
NNK_translation	BAPV_translation	merged_translations
NaN	R*IATPCVETPLFL	R*IATPCVETPLFL
NaN	CPTAT*LQLTLCGE	CPTAT*LQLTLCGE
NaN	LFIATRT*VTSNA*	LFIATRT*VTSNA*
ASIQCP*RGCPAHY	ASIQCP*RGCPAHY	ASIQCP*RGCPAHY

Prototype workflow

Joining Paired-End Reads



Isolation of Randomized Insert



Translation

11. The Bacterial, Archaeal and Plant Plastid Code

TTT F Phe	TCT S Ser	TAT Y Tyr	TGT C Cys
TTC F Phe	TCC S Ser	TAC Y Tyr	TGC C Cys
TTA L Leu	TCA S Ser	TAA * Ter	TGA * Ter
TTG L Leu i	TCG S Ser	TAG * Ter	TGG W Trp
CTT L Leu	CCT P Pro	CAT H His	CGT R Arg
CTC L Leu	CCC P Pro	CAC H His	CGC R Arg
CTA L Leu	CCA P Pro	CAA Q Gln	CGA R Arg
CTG L Leu i	CCG P Pro	CAG Q Gln	CGG R Arg
ATT I Ile i	ACT T Thr	AAT N Asn	AGT S Ser
ATC I Ile i	ACC T Thr	AAC N Asn	AGC S Ser
ATA I Ile i	ACA T Thr	AAA K Lys	AGA R Arg
ATG M Met i	ACG T Thr	AAG K Lys	AGG R Arg

Sequence data basis

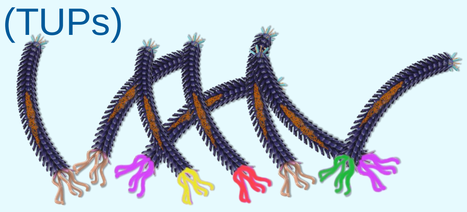
NNK_translation	BAPV_translation	merged_translations	NNK_type
EEHKCKKLMCLLTK	EEHKCKKLMCLLTK	EEHKCKKLMCLLTK	TRUE
	LFSLTRPLKT*VVH	LFSLTRPLKT*VVH	FALSE
	GL*QTD*CDTHSFF	GL*QTD*CDTHSFF	FALSE
DERRCGVQ*CWQMT	DERRCGVQ*CWQMT	DERRCGVQ*CWQMT	TRUE
	LLASTPQVITTVY*	LLASTPQVITTVY*	FALSE
	GL*QTD*CDTHSFF	GL*QTD*CDTHSFF	FALSE
RSR*CQHNYCRSTD	RSR*CQHNYCRSTD	RSR*CQHNYCRSTD	TRUE
*TVQCMRLMCSIAF	*TVQCMRLMCSIAF	*TVQCMRLMCSIAF	TRUE

vectorized data frames for faster data query

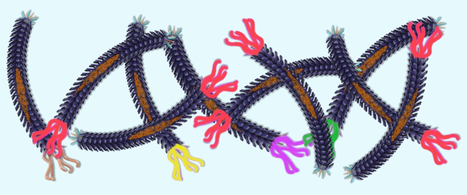
To Separate the Indistinguishable

- Difficult to distinguish between true binders and Target Unrelated Peptides (TUPs)

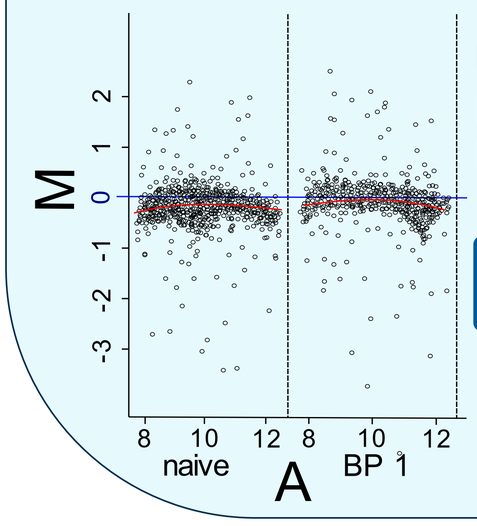
1. Round



2. Round



- Gradual increase of frequency in fast-propagating phage clones



Find fast propagating population

Statistical Methods to Discover Binding Peptides

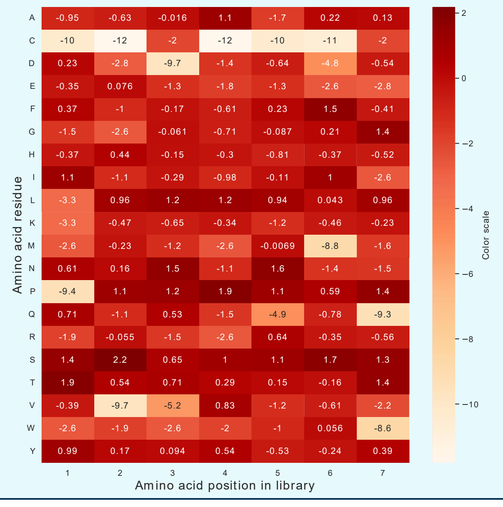
- Identifying TUPs with differential enrichment analysis[1]

Trimmed mean of M-values normalization and PCR

Testing of best fitting distributions and models

Checking false discovery rate and running statistical tests

- Identifying binding motifs with a derivative of the PepSimili[2] approach



Compare peptide motifs with protein domains using metal binding protein data base

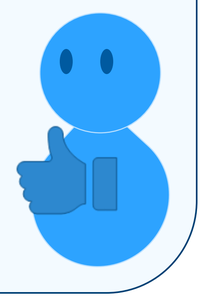
Future Outlook & Conclusion

Check optimal normalization methods

Consider data origin & experimental conditions !

Best tests for a multi-trail Phage display experiment !

Phage display trails are a complex evolution theory task with no current easy established solution to find best binders for inorganic materials



This research was funded by the German Federal Ministry of Education and Research (BMBF), [grant no. 031B1348A]
Helmholtz Institute Freiberg for Resource Technology · Department of Biotechnology
M.Sc. Christoph Bloß c.bloss@hzdr.de
[1] Robinson, M.D., Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11, R25 (2010). https://doi.org/10.1186/gb-2010-11-3-r25
[2] Vekris A, Pilalis E, Chatziioannou A and Petry KG (2019) A Computational Pipeline for the Extraction of Actionable Biological Information From NGS-Phage Display Experiments. Front. Physiol. 10:1160. doi: 10.3389/fphys.2019.01160

SPONSORED BY THE Federal Ministry of Education and Research