Highly repetitive elements hold promise as a rich source of cancer vaccine targets

Authors: Wilson McKerrow[1], Simon Chu[1], Nicole Van Hoeken[1], Megan Snyder[1], Liyang Diao[1], Heike Keilhack[1]

Affiliations: [1]ROME Therapeutics, Boston MA, USA

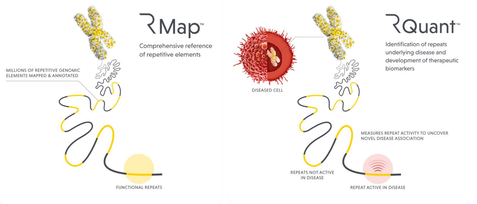
Contact: wmckerrow@rometx.com

INTRODUCTION

BACKGROUND: Approximately half of the human genome is derived from repetitive sequences that can copy themselves to new genomic loci. Typically silenced in healthy somatic tissues but de-repressed in tumors, expression of these elements, referred to as the "repeatome", has been linked to inflammation and genome instability. Peptides derived from repetitive elements have been discussed as potential cancer vaccine candidates, with the added benefit that immune escape cannot be achieved through the silencing of a single locus. However, accurate identification of which repeat transcripts are expressed in a tumor specific manner has remained a challenge.

ROME'S DATA SCIENCE PLATFORM ADDRESSES DARK GENOME CHALLENGES

ROME Therapeutics has developed a first-in-class repeat quantification platform built specifically to address challenges in this space



CHALLENGE: Vastness
Difficult to control false positives

ROME Tx solution
- Comprehensive repeatome database
- Focus on the parts of the dark genome with the potential to yield protein products

CHALLENGE: Repetitiveness
Difficult to understand what's actually being expressed

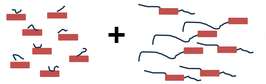
ROME Tx solution
- Probabilistic modeling / machine learning
- Specifically quantify those transcripts that are likely to be translated
- Validate translation in proteomic data

Canonical repeat proteins come from highly repetitive loci that can only be quantified using specialized tools

Sources of noise

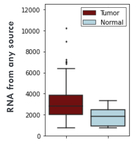
DNA contamination / transcripts amplified across 1000s of repetitive loci

Untranslated copies embedded in introns and 3' UTRs



Standard quantification

High repeat expression "hallucinated"



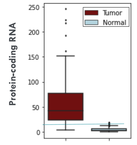
Extensive manual curation of repeats + custom algorithms

Curated database of repeat promoters and ORFs

Repeat aware alignment strategies

RQuant

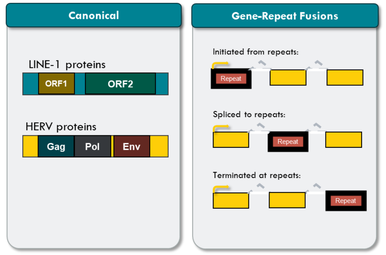
Tumor specificity of repeat targets revealed



DARK GENOME TARGETS ARE "NEEDLES IN THE HAYSTACK"

ROME is targeting multiple classes of repeat antigen

Focus of this poster | See poster #1204



Millions of repetitive loci

RMap™

1,000s of loci with protein expression potential

MHC profiling

10,000s of peptides with predicted MHC binding

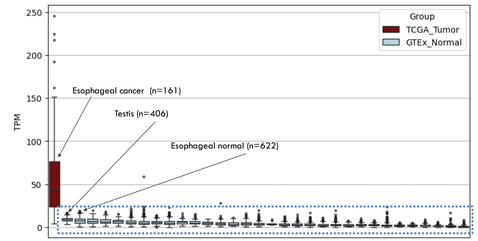
RQuant™

100s of peptides that are overexpressed in cancer

A PROMISING TARGET IN ESOPHAGEAL CANCER

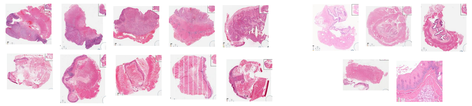
REPEAT PROTEINS HAVE HIGH TUMOR SPECIFICITY

ROME data science platform has identified a family of repeat proteins-highly expressed in esophageal cancer (TCGA, GTEx data)



Protein level data needed to determine whether the lower RNA signal in normal is background or evidence of translation

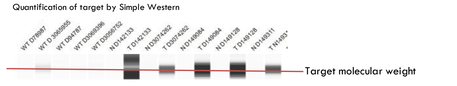
Esophageal cancer and normal samples were acquired to assess tumor specificity of the target at the protein level

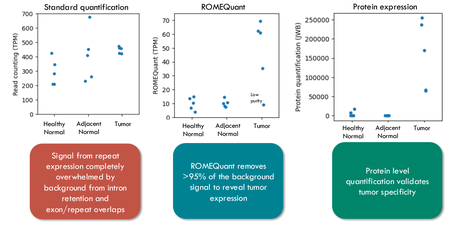


- 5 Matched Tumor/Normal Pairs (frozen)
- Esophageal squamous cell carcinoma
- Vietnam, male
- 5 Normal esophagus (frozen)
- No cancer
- USA, 4 female / 1 male
- Data generated: RNA-seq, Simple Western protein quantification, immunopeptidomics

Protein level data validates tumor specificity of this target

Quantification of target by Simple Western





Signal from repeat expression completely overwhelmed by background from intron retention and exon/repeat overlaps

ROMEQuant removes >95% of the background signal to reveal tumor expression

Protein level quantification validates tumor specificity

REPEAT PEPTIDES ARE PRESENTED BY DIVERSE CLASS I MHC

Class I MHC presentation in esophageal cancer (publicly available data, European descent)

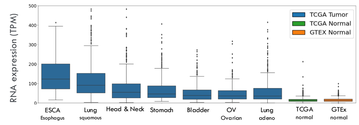
	Tumor 1	Tumor 2	Tumor 3	Tumor 4	Tumor 5	Tumor 6	Tumor 7	Consensus
Peptide A	X							Yes
Peptide B							X	No
Peptide C				X				Yes
Peptide D					X			Yes
Peptide E						X		Yes
HLA-A	A03:01	A02:01, A29:02	A01:01, A11:01	A01:01, A24:01	A23:01, A33:01	A01:01, A23:01	A02:01	
HLA-B	B07:02, B08	B15:01, B18:01	B08:01, B35:01	B08:01, B15:01	B14:02, B49:01	B08:01, B44:03	B44:02, B58:01	

Data from Nicholas et al., 2023, Identification of neoantigens in esophageal adenocarcinoma, Immunology.

Class I MHC presentation also observed in ROME samples after adjusting for population specific sequence variation

Peptide	Tumor 1	Tumor 2	Tumor 3	Tumor 4	Tumor 5	Consensus	Population specific
Peptide F			X	X		No	No
Peptide G				X		No	Yes
Peptide H		X				No	Yes
HLA-A	A02:07	A24:07	A11:303, A02:07	A02:03	A11:303, A11:01		
HLA-B	B46:01	B58:01, B27:06	B46:01, B51:06	B38:02, B55:02	B15:25, B54:01		

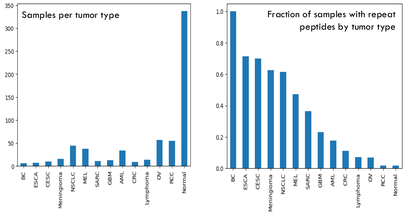
Target is expressed in additional tumor types



Presentation of repeat peptides predicted by RNA-seq is validated in immunopeptidomics across tumor types

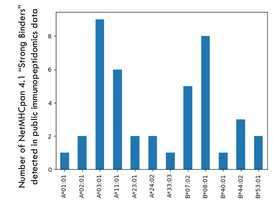
Immunopeptidomics data collected from 40 studies in the public domain, including:

- Tumor (~1200 MS runs)
- Normal (~1000 MS runs)
- Cell line (~800 MS runs)

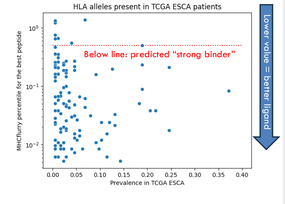


REPEAT PEPTIDES HAVE POTENTIAL AS A POPULATION VACCINE

Repeat peptides span a range of sequences and HLA alleles



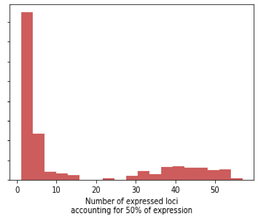
(Showing alleles present in ≥15% of North American White, Black, Hispanic or Asian population)



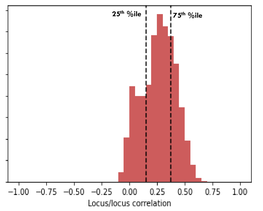
94% of alleles in TCGA ESCA predicted to present at least 1 of the peptides detected in our reanalysis of public immunopeptidomics data

DUE TO THEIR HIGHLY REPETITIVE NATURE, IMMUNE ESCAPE FROM THERAPIES TARGETING REPEATS MAY BE LESS LIKELY

Number of loci expressing each peptide



Correlation between loci expressing these antigens



Many peptides are encoded by dozens of expressed transcripts that are only partially co-regulated

CONCLUSION

SUMMARY: Many repetitive elements are specifically expressed in tumors, making them promising cancer vaccine targets. Furthermore, immune escape may be less straightforward for an antigen that is expressed from multiple loci. However, their repetitiveness makes their expression difficult to quantify. ROME has developed tools that make it possible to accurately quantify repeat element expression and identify promising immunotherapy targets derived from repetitive elements. Here, we highlight a family of closely related repeat proteins that are expressed and presented in several tumor types, most especially esophageal cancer. We were able to validate that this target is tumor specific at the protein level. Based on HLA analysis, we believe this target could form the basis of a population vaccine and perhaps even be the target of a prophylactic cancer vaccine.