A. COVER PAGE

Project Title: Prioritizing rare variants associated with cancer using non-coding annotation	
Grant Number: 5R01HG008126-02	**Project/Grant Period:** 07/01/2016 - 06/30/2019
Reporting Period: 07/01/2016 - 06/30/2017	**Requested Budget Period:** 07/01/2017 - 06/30/2018
Report Term Frequency: Annual	**Date Submitted:**
Program Director/Principal Investigator Information: MARK BENDER GERSTEIN , PHD AB **Phone number:** 203- 432-6105 **Email:** mark.gerstein@yale.edu	**Recipient Organization:** YALE UNIVERSITY YALE UNIVERSITY OFFICE OF SPONSORED PROJECTS PO BOX 208327 NEW HAVEN, CT 065208327 **DUNS:** 043207562 **EIN:** 1060646973A1 **RECIPIENT ID:**
Change of Contact PD/PI: N/A	
Administrative Official: SUSAN HEDLEY Office of Sponsored Projects P.O. Box 208327 New Haven, CT 065208327 **Phone number:** 12037854689 **Email:** OSP@YALE.EDU	**Signing Official:** SUSAN HEDLEY Office of Sponsored Projects P.O. Box 208327 New Haven, CT 065208327 **Phone number:** 12037854689 **Email:** OSP@YALE.EDU
Human Subjects: No	**Vertebrate Animals:** No
hESC: No	**Inventions/Patents:** No

B. ACCOMPLISHMENTS

B.1 WHAT ARE THE MAJOR GOALS OF THE PROJECT?

Aim 1. Adapt our existing tool for prioritizing somatic variants (FunSeq) to create a generalizable approach for prioritizing impactful non-coding variants (eleVAR).

Aim 2. Implement eleVAR pipeline and develop a workflow for tuning and assessing performance, focusing on prostate cancer as a test case for a specific disease.

Aim 3. High-throughput experimental characterization of ~1200 variants using Clone-seq and luciferase reporter assays.

Aim 4. Detailed experimental validation of a few non-coding variants from eleVAR.

B.1.a Have the major goals changed since the initial competing award or previous report?

No

B.2 WHAT WAS ACCOMPLISHED UNDER THESE GOALS?

File uploaded: Accomplishments.pdf

B.3 COMPETITIVE REVISIONS/ADMINISTRATIVE SUPPLEMENTS

For this reporting period, is there one or more Revision/Supplement associated with this award for which reporting is required?

No

B.4 WHAT OPPORTUNITIES FOR TRAINING AND PROFESSIONAL DEVELOPMENT HAS THE PROJECT PROVIDED?

File uploaded: Individual Developement Plan.pdf

B.5 HOW HAVE THE RESULTS BEEN DISSEMINATED TO COMMUNITIES OF INTEREST?

NOTHING TO REPORT

B.6 WHAT DO YOU PLAN TO DO DURING THE NEXT REPORTING PERIOD TO ACCOMPLISH THE GOALS?

* Scaling up cloning pipeline after initial small-scale trial to clone and test the reporter activity of 200 WT predicted enhancer elements.
* Testing the reporter activity of 2 variants per WT enhancer element generated using Clone-seq to assess the success of predictions.
* Performing large-scale target mutagenesis to introduce two SNVs per element using our Clone-seq pipeline.
* Tuning parameters of our statistical model using large-scale experimental results.

We initiated the project in June 2016 and started to collect data and information. During this period, the Gerstein, Rubin and Yu labs discussed the collaboration details and made a plan for a regular bi-weekly working group call. The first working group call was held on Sept 12, with the Gerstein lab reporting an update in statistical modeling. This was followed by regular calls. On Feb 20, 2017, the Yu and Rubin labs visited Yale University for a two-day meeting.

In related to Aim1: The Gerstein Lab built a statistical model to predict and prioritize germline variants that can significantly affect enhancer regulatory activity, which employs conservation score, histone modification, transcription factor binding profile, CAGE data information. The model can do both classification and regression: the classification model can predict a probability whether variants can significantly change the regulatory activity while the regression model can predict the regulatory activity change between mutant and wild type (fold change). To iteratively tune parameter using experimental results, we developed an online learning framework to refine model for **Aim 2** and then select about 200 regions and predicted variants with different level of regulatory changes to further tune parameters for **Aim 3**.

Besides model learning, we have conducted a case study to identify and prioritize impactful noncoding variants on Renal cell carcinoma, which accounts for more than 90% of kidney cancers. Papillary renal cell carcinoma (pRCC) is the second most common subtype of renal cell carcinoma. We carry out the first whole-genome study of pRCC to discover triggering DNA changes explaining these cases. We find two impactful noncoding variants which are associated with prognosis and can aid clinical decisions. The results have been published PLOS Genetics on Mar 30, 2017.

In related to Aim3: The Rubin lab assessed experimental conditions and developed a high-throughput pipeline for WT element cloning into a pDEST-hSCP1-luc luciferase reporter vector on a small selection (14) of predicted enhancer elements. Elements were amplified via PCR with primers containing attB1 and attB2 sequences and cloned into pDONR223 using Gateway BP clonase. Four colonies for each element were picked and sequenced via Sanger sequencing. 54 out of the 56 colonies contained the precisely cloned WT element with no nucleotide changes, indicating a PCR-induced error rate of <4%.

One clone for each element with the correct sequence was then cloned into pDEST-hSCP1-luc using Gateway LR clonase, and luciferase reporters containing the elements were then transfected into K562 cells. 13 of the 14 predicted enhancer elements demonstrated positive enhancer activity indicating a high success rate for our predictions.

To construct a positive control set for the enhancer activity assays, the Yu Lab cloned several widely used promoters that have been implied to possess enhancer activities. These include CMV (cytomegalovirus) promoter, PGK (phosphoglycerate kinase) promoter, SV40 (Simian virus 40) promoter, and RSV (Rous sarcoma virus) promoter. They delivered the luciferase assay vector containing the promoters into K562 cells via electroporation and detected up to 9-fold higher enhancer activities in these elements compared to the CDS of EGFP by dual luciferase assay.

Based on the prediction, they used a webtool to design site-directed mutagenesis primers to introduce the target SNVs into the 14 elements. Two SNVs will be introduced to each element respectively to generate two single-SNV-containing alleles. We predicted one of the SNVs would result in a significant change in the enhancer activity while the other would not. The experiments for mutagenesis and the following enhancer activity comparison between the WT and mutant elements are currently in progress.

In summary, we mainly focus on the data mining and experimental setup in the first year. The most challenging work is to build a quantitative model, screen reliable enhancer regions and predict potentially effective variants for experimental validation.

Training and professional development

All graduate students supported by an NIH award at Yale University are required to create an individual development plan. Students provide updates on their IDP activities as part of their annual thesis committee meetings, and documentation is retained by the students' graduate programs.

All postdoctoral trainees at Yale University are required to create an individual development plan and to provide annual progress reports for review by, and discussion with, the faculty mentor. Progress reports are then submitted to the Yale Office for Postdoctoral Affairs as a condition of the trainee's reappointment by this Office.

C. PRODUCTS

C.1 PUBLICATIONS

Are there publications or manuscripts accepted for publication in a journal or other publication (e.g., book, one-time publication, monograph) during the reporting period resulting directly from this award?

Yes

Publications Reported for this Reporting Period

Public Access Compliance	Citation
Complete	Chen J, Wang B, Regan L, Gerstein M. Intensification: A Resource for Amplifying Population-Genetic Signals with Protein Repeats. Journal of molecular biology. 2017 February 3;429(3):435-445. PubMed PMID: 27939289.
Complete	Li S, Shuch BM, Gerstein MB. Whole-genome analysis of papillary kidney cancer finds significant noncoding alterations. PLoS genetics. 2017 March;13(3):e1006685. PubMed PMID: 28358873; PubMed Central PMCID: PMC5391127.

C.2 WEBSITE(S) OR OTHER INTERNET SITE(S)

Nothing to report

C.3 TECHNOLOGIES OR TECHNIQUES

NOTHING TO REPORT

C.4 INVENTIONS, PATENT APPLICATIONS, AND/OR LICENSES

Have inventions, patent applications and/or licenses resulted from the award during the reporting period?

No

C.5 OTHER PRODUCTS AND RESOURCE SHARING

NOTHING TO REPORT

D. PARTICIPANTS

D.1 WHAT INDIVIDUALS HAVE WORKED ON THE PROJECT?

Commons ID	S/K	Name	Degree(s)	Role	Cal	Aca	Sum	Foreign Org	Country	SS
MGERSTEIN	Y	Gerstein, Mark Bender	AB,PHD	PD/PI	0	1	0			NA
GGURSO2	N	Gursoy, Gamze	PHD,BS	Postdoctoral Scholar, Fellow, or Other Postdoctoral Position	4	0	0			NA
MAR123	Y	RUBIN, MARK A.	MD	Co-Investigator	1	0	0			NA
HAIYUANYU	Y	Yu, Haiyuan	PHD	Co-Investigator	0	0	1			NA
JINLIANG	N	Liang, Jin		Postdoctoral Scholar, Fellow, or Other Postdoctoral Position	3	0	0			NA
KUMARY	N	Kumar, Yugandhar		Postdoctoral Scholar, Fellow, or Other Postdoctoral Position	2	0	0			NA
RFRAGOZA	N	Fragoza, Robert	BS,PHD	Graduate Student (research assistant)	6	0	0			NA
TANGFANYING	N	Tang, Fanying	PHD,BS	Graduate Student (research assistant)	12	0	0			NA
TINGYIWANG	N	Wang, Ting-Yi	BS,MS	Postdoctoral Scholar, Fellow, or Other Postdoctoral Position	2	0	0			NA
EKTAKHURANA	N	KHURANA, EKTA	PHD	Assistant Professor	1	0	0			NA
KOONKIUYAN	N	YAN, KOON-KIU	PHD	Postdoctoral Scholar, Fellow, or Other Postdoctoral Position	2	0	0			NA
DICHAKRAVARTY	N	Chakravarty, Dimple	PHD	Postdoctoral Scholar, Fellow, or Other Postdoctoral Position	4	0	0			NA

PEMANI	N	Emani, Prashant Siva	BS,MS,PHD	Postdoctoral Scholar, Fellow, or Other Postdoctoral Position	2	0	0			NA

Glossary of acronyms:
S/K - Senior/Key
DOB - Date of Birth
Cal - Person Months (Calendar)
Aca - Person Months (Academic)
Sum - Person Months (Summer)
Foreign Org - Foreign Organization Affiliation
SS - Supplement Support
RE - Reentry Supplement
DI - Diversity Supplement
OT - Other
NA - Not Applicable

D.2 PERSONNEL UPDATES

D.2.a Level of Effort

Will there be, in the next budget period, either (1) a reduction of 25% or more in the level of effort from what was approved by the agency for the PD/PI(s) or other senior/key personnel designated in the Notice of Award, or (2) a reduction in the level of effort below the minimum amount of effort required by the Notice of Award?

No

D.2.b New Senior/Key Personnel

Are there, or will there be, new senior/key personnel?

No

D.2.c Changes in Other Support

Has there been a change in the active other support of senior/key personnel since the last reporting period?

Yes

File uploaded: NIH_OS_0517.pdf

D.2.d New Other Significant Contributors

Are there, or will there be, new other significant contributors?

No

D.2.e Multi-PI (MPI) Leadership Plan

Will there be a change in the MPI Leadership Plan for the next budget period?

NA

E. IMPACT

E.1 WHAT IS THE IMPACT ON THE DEVELOPMENT OF HUMAN RESOURCES?

Not Applicable

E.2 WHAT IS THE IMPACT ON PHYSICAL, INSTITUTIONAL, OR INFORMATION RESOURCES THAT FORM INFRASTRUCTURE?

NOTHING TO REPORT

E.3 WHAT IS THE IMPACT ON TECHNOLOGY TRANSFER?

Not Applicable

E.4 WHAT DOLLAR AMOUNT OF THE AWARD'S BUDGET IS BEING SPENT IN FOREIGN COUNTRY(IES)?

Dollar Amount	Country
13695	ITALY

F. CHANGES

F.1 CHANGES IN APPROACH AND REASONS FOR CHANGE

Not Applicable

F.2 ACTUAL OR ANTICIPATED CHALLENGES OR DELAYS AND ACTIONS OR PLANS TO RESOLVE THEM

NOTHING TO REPORT

F.3 SIGNIFICANT CHANGES TO HUMAN SUBJECTS, VERTEBRATE ANIMALS, BIOHAZARDS, AND/OR SELECT AGENTS

F.3.a Human Subjects

No Change

F.3.b Vertebrate Animals

No Change

F.3.c Biohazards

No Change

F.3.d Select Agents

No Change

G. SPECIAL REPORTING REQUIREMENTS

G.1 SPECIAL NOTICE OF AWARD TERMS AND FUNDING OPPORTUNITIES ANNOUNCEMENT REPORTING REQUIREMENTS

NOTHING TO REPORT

G.2 RESPONSIBLE CONDUCT OF RESEARCH

Not Applicable

G.3 MENTOR'S REPORT OR SPONSOR COMMENTS

Not Applicable

G.4 HUMAN SUBJECTS

G.4.a Does the project involve human subjects?

No

G.4.b Inclusion Enrollment Data

Not Applicable

G.4.c ClinicalTrials.gov

Does this project include one or more applicable clinical trials that must be registered in ClinicalTrials.gov under FDAAA?

G.5 HUMAN SUBJECTS EDUCATION REQUIREMENT

Are there personnel on this project who are newly involved in the design or conduct of human subjects research?

G.6 HUMAN EMBRYONIC STEM CELLS (HESCS)

Does this project involve human embryonic stem cells (only hESC lines listed as approved in the NIH Registry may be used in NIH funded research)?

No

G.7 VERTEBRATE ANIMALS

Does this project involve vertebrate animals?

No

G.8 PROJECT/PERFORMANCE SITES

Organization Name:	DUNS	Congressional District	Address
Primary: Yale University	043207562	CT-003	Yale University 266 Whitney Avenue, BASS 432A New Haven CT 065208114
Weill Medical College of Cornell University	060217502	NY-012	1300 York Avenue, box 69 New York NY 100654805
Cornell University	872612445	NY-023	341 Pine Tree Road Ithaca NY 148502820
Universita degli Studi di	436121107		Via Sommarive 14

Trento			38123 Povo Trento

G.9 FOREIGN COMPONENT

Organization Name: Universita degli Studi di Trento
Country: ITALY
Description of Foreign Component:
No co-authorship based on this study has occurred at this time. The facilities at University of Trento have been used in this study as described in the statement of work, providing an environment to analyze data sets of lareg scale cohorts. No financial support has been received from University of Trento for this study.

G.10 ESTIMATED UNOBLIGATED BALANCE

G.10.a Is it anticipated that an estimated unobligated balance (including prior year carryover) will be greater than 25% of the current year's total approved budget?

No

G.11 PROGRAM INCOME

Is program income anticipated during the next budget period?

No

G.12 F&A COSTS

Is there a change in performance sites that will affect F&A costs?

No