CENTER FOR DATA INSIGHTS AND INNOVATION
CALIFORNIA HEALTH & HUMAN SERVICES AGENCY

View xForm - Project Application v6

This form is for new projects that have not been previously approved by CPHS.

Data entry

Amendment Header

Amendment Submitter

August 2024 cycle

_____Full board Amendment_____

06/13/2024 • Nicholas Zadrozna • Internal

IMPORTANT NOTE FROM THE PRIMARY REVIEWER OF THIS PROJECT:

1. In the description of amendment changes, would you please clearly describe all changes you plan to make.

2. Throughout the body of the application, when you describe changes, would you please leave the original language unchanged and describe the changes at the end of the original text introduced with the note " June 2024 amendment update:". In this way I will be able to understand what you plan to do differently.

3. Please provide additional consent forms and any other changed forms or documents that will be used in the new amended version of the study.

06/18/2024 • Sussan Atifeh • *Not* Internal

A COPY OF THE EMAIL FROM THE PRIMARY REVIEWER OF THIS PROJECT TO THE RESEARCH TEAM:

Good morning Dr. Erhunmwunsee—

I am the reviewer for the amendment submitted for the project referenced above.

I have reviewed the amendment and I am having a little trouble understanding exactly what you plan to do differently.

You have indicated that you are changing recruitment procedures, research procedures, and research questions. Your description of the changes talks about changing the consent procedures, but nothing else—are there changes to the research procedures and/or research questions that haven't been described? If this is moving from the pilot study

phase to another phase do you anticipate increasing the original sample size?

I see that there are many areas in the application, such as description of the research procedures, where you have changed the text, but because of the complexity of your study it is difficult for me to understand what new language is specific to this amendment.

I also did not find any new consent forms. Given that you are changing your consent procedures I am assuming that there are concomitant changes to the consent form(s)?

To assist in the review process, would you please revise the amendment in the following ways—

1. In the description of amendment changes, would you please clearly describe all changes you plan to make.
2. Throughout the body of the application, when you describe changes, would you please leave the original language unchanged and describe the changes at the end of the original text introduced with the note " June 2024 amendment update:". In this way I will be able to understand what you plan to do differently.
3. Please provide additional consent forms and any other changed forms or documents that will be used in the new amended version of the study.

Please be aware that, because your changes involve changes in the human subject contact aspects of your study, it may be necessary to have this amendment reviewed by the full CPHS board. I will be able to let you know whether this will be necessary after I have a better understanding of all the changes you are proposing, especially your revised consent procedures and consent forms.

Thank you.

Laura

Laura E. Lund, MA
Member
Committee for the Protection of Human Subjects

06/18/2024 • Sussan Atifeh • *Not* Internal

Danielle Shores, B.A
Email: dshores@coh.org **Business:** (626) 218-0091

Instructions for amending your approved application:

This is a copy of the project application in order to amend the project. You must answer all the amendment questions. After you've answered those questions, you will have to update all answers on the form that related to your proposed changes. You may leave other questions with their original answer. If you do not update the appropriate responses on the form related to your proposed amendment, you will be required to make additional changes.

Note that the contacts listed on this page are output only questions that cannot be changed. If you need to request personnel changes, you will be prompted later on within this form to enter the new contact information.

PI:

Loretta Erhunmwunsee, MD
Email: lorettae@coh.org **Business:** (626) 218-7287

Administrative Contacts:

Name	Role

Protocol Number:

2022-004

Protocol Title:

The Impact of Racism-Related Socio-Environmental Factors on African-American Non-Small Cell Lung Cancer (NSCLC) Mutational Signatures

Indicate what types of changes you are requesting to this project. Select all that apply

Please re-check your choices in this section and ensure you have selected relevant options.

06/18/2024 • Sussan Atifeh • *Not* Internal

Recruitment strategy and/or materials
Population, sample size, inclusion/exclusion criteria

Clearly summarize and justify your proposed changes to the protocol in layman's terms for all selections made above

In this section please clearly describe all changes you plan to make.

06/18/2024 • Sussan Atifeh • *Not* Internal

If the proposed promotional materials/website content are not available yet it would be best to remove references to your plan to use promotional materials and a website for recruitment from the current amendment. This change to your recruitment procedures can be submitted as a separate amendment at a later time when the materials are available for review.

07/17/2024 • Laura Lund, MA • *Not* Internal

If PHI project personnel will be accessing the CCR data please name the project personnel being added.

07/17/2024 • Laura Lund, MA • *Not* Internal

In the project procedures, would you please describe PHI's relationship to this study? Generally speaking, the CCR is contracted through PHI, so if PHI's only role in this study is data delivery as the CCR contractor then they (and PHI staff) do not need to be listed as a separate project site and the PHI personnel preparing the data for delivery do not need to be listed as project personnel. If PHI's role goes beyond data delivery as the CCR contractor and they are otherwise involved in this project then this role should be clarified.

07/17/2024 • Laura Lund, MA • *Not* Internal

This was a pilot study to obtain data for an NIH R01 research grant, which was recently awarded as an R37. We are changing the aims, increasing our sample size to N=100, and extending our cohort dates for this study to align directly with the R37. During our expansion, we would also like to add a new site, Public Health Institute (PHI). Please read below for further details on all proposed changes.

Additional Recruitment Site: We would like to add another site, Public Health Institute (PHI), during our study's expansion. Public Health Institute's Cancer Registry of Greater California (PHI/CRGC) program would generate a list of eligible patients from (CRGC) and send it to City of Hope for our study staff to contact the potential participants. PHI will serve as a data repository, in that they will be sending us contact information and COH staff will reach

out to the patients. PHI staff will not be accessing CCR data, as all recruitment and enrollment will be taking place by COH research staff. We are requesting some data from CRGC (please see CCR support letter attached): names, contact information, sex, age, disease stage, and NSCLC surgical procedure. City of Hope staff will enroll patients in the study by conducting the consent process and collecting surveys. Completed informed consent and HIPAA authorization will be provided to PHI staff who will request tissue samples from CRGC and send them to COH for further analysis. PHI will serve as a data repository and specimen delivery service. We have listed them as a separate entities and have listed PHI as a separate due to the way we will receive specimen. COH will receive specimen from participants recruited from CRGC through PHI repository staff, whereas participants from LACR will be received through the USC repository system. The reason why PHI is listed seperately, and not within the existing name in the original submission is due to the which entity will be sending us the specimen.

Sample Size Increase: We will be expanding this preliminary study by increasing the sample size for our California cohort to N=100 to align with our recently awarded R37.

Recruitment Strategy/Material: We would also like to streamline our consent process and introduce promotional materials to increase recruitment. During this expansion, we have partnered with a site who successfully carried out a similar study with African American males, using a one-part consent process for the conduction of surveys and whole genome sequencing of prostate tumor tissue samples. They gave patients the option to complete the informed consent form, HIPAA authorization, and questionnaire with guidance or independently. This significantly increased their accrual, and we would like to model their successful approach. Therefore, we are requesting to waive an element of the informed consent process. We are proposing that patients have the option to complete the Informed Consent Form and HIPPA Authorization independently with written step-by-step guidance on the cover page of the consent sheet (Please see attached "Cover letter and Form Instruction"). Patients will also have access to a QR-linked video guide for filling out the consent form and still have the option to contact City of Hope study staff to be guided through the packet, as stated in the consent form. We have provided the video script for reference (Please see attached "Video Consent Script").

We would also like to combine our consent process into a one-part consent. Currently, we are calling patients two separate times to consent them for Part 1 (the questionnaire) and Part 2 (medical record release and whole genome sequencing) of our study. This has led to ~40% of patients lost-to-follow up after completing the consent for Part 1. After receiving patient feedback, we believe having two separate consents places an undue burden on our minoritized communities. Therefore, we would like to combine this into a one-part consent to decrease the number of touchpoints with participants. Using this method, patients will only have to review and complete one informed consent for both parts of the study and will be fully enrolled in the study after returning the signed consent form. Patients would receive a $75 Visa gift card ($25 for completing the questionnaire and $50

for completing the medical release form) upon receiving the completed informed consent, survey, and HIPAA Medical Record Release.

We are also proposing to implement promotional materials, such as a website, for patients to gain more access to our study. No patient PHI or personal information will be used for this website, as the purpose is to engage our participants and encourage genuine interest in our study. This website is still in development. Materials that are not approved by CPHS will not be used prior to approval. Per COH IRB, the usage of promotional material needs to be approved before development. After COH IRB gives approval for development, promotional materials will be curated towards study activities and then submitted to the both CPHS and COH IRBs for approval to disseminate to potential or enrolled participants.

Indicate the Level of Risk involved with the changes proposed.

If level of risk has changed, please update the "Risks" section in the protocol form.

Level of Risk has not changed

PI City Output *(Internal)*

Duarte

PI Location State Output *(Internal)*

California

Personnel Information for Amendment

Please complete the questions below.

If while trying to complete those questions, personnel are not found by their email address, you can add them in the system by completing the 'new contact form'. Click on the form and complete it. Within a few minutes of completing the form you will receive an email notifying you of the availability of the new contact. You should then be able to add them in the subsequent questions.

New Contact Form

Existing Personnel

If new PHI personnel will have access to the data please provide their names in the amendment description above. (Please see the note above for clarification on whether this is necessary.)

07/17/2024 • Laura Lund, MA • *Not* Internal

Name	Role
Aamna Akhtar, BS	Research Team
Christina Chang	Research Team
Danielle Shores, B.A	Research Team
Loretta Erhunmwunsee, MD	Research Team
Loretta Erhunmwunsee, MD	Principal Investigator
Nicole Herrara, BA	Responsible Official
Peter Vien, Bachelor's of Science	Research Team
Yi Xiao, Master in Science	Research Team

Will you be making any changes to the makeup of research personnel?

The amendment description states that you will be adding personnel.

07/17/2024 • Laura Lund, MA • *Not* Internal

*No change in personnel

Project Information

SUBMITTER

Application completed by:

Danielle Shores, B.A

Email: dshores@coh.org **Business:** (626) 218-0091

PREVIOUSLY APPROVED EXEMPTION

Is there a previously-approved exemption from CPHS for this project?

No

PROJECT TITLE

Enter the project title (please capitalize each word in your title).

The Impact of Racism-Related Socio-environmental factors on African-American NSCLC mutational signatures

STUDY PROCEDURES

Indicate the study procedures involved in this research. Check all that apply.

Recruitment-Participant
Specimen Registry
Surveillance Data
Surveys

TYPE OF RESEARCH REQUEST

Indicate which of the following applies to this research. Check all that apply.

*Death Data Only refers to health-related studies requesting existing mortality data from **within** the California Human Health Services Agency (CHHSA)*

*SB-13 (Information Practices Act) refers to health-related studies requesting existing data from **outside** the CHHSA (e.g. California Department of Corrections and Rehabilitation [CDCR], California Department of Education [CDE], etc.) **OR** studies requesting data **within** the CHHSA that are not state funded or involving state staff.*

Common Rule/Human Subjects refers to health-related studies that involve direct or indirect interaction with human subjects (e.g. recruitment, interviews, etc.)

*Common Rule Only refers to health-related studies requesting existing data from **within** the CHHSA (e.g. Office of Statewide Health Planning and Development [OSHPD], California Department of Public Health [CDPH], etc)*

Common rule/Human subjects

PROJECT TYPE DETAILS

Indicate which, if any, apply to this research. Check all that apply.

If the research does not involve any of following, choose "None of the above."

Minimal Risk
HIPAA waiver
Consent form

VULNERABLE POPULATIONS

Indicate which vulnerable populations, if any, will be involved with this research. Check all that apply.

If vulnerable populations are not part of the research, choose "Not applicable."
Note regarding minors: in the United States, a minor is under 18 years of age. If research is conducted outside the United States, a minor is under the age of majority in the countries where research is to be conducted.

Not applicable

FUNDING

Is this research funded?

Yes

Indicate the funding source for this project.
Federally funded

Enter name of federally-funded source.
National Institutes of Health

EXPEDITED REVIEW CONSIDERATION

Please check the criteria below that you think your project meets to qualify for an expedited review. If none of these expedited criteria are appropriate for your project, choose 'not applicable'; your protocol will be reviewed by the full committee. Note that CPHS will make the final determination of whether the project meets the criteria for expedited review.

Since this section is related to the original project, please re-select the original option, "Not applicable."

06/18/2024 • Sussan Atifeh • *Not* Internal

Protected Health Information/Personally Identifiable Data (PHI/PID) is defined as information in any format that identifies the individual, including demographic information collected from an individual that can reasonably be used to identify the individual. Additionally, PHI is information created or received by a healthcare provider, health plan, employer, or health care clearinghouse; and relates to the past, present, or future physical or mental health or condition of an individual, including any of the 18 HIPAA identifiers.

Note: Please be aware that individual participants may be identifiable by combining other items in the data even when none of the 18 HIPAA identifiers are present. Thus, a study may still contain PID even after removing or never acquiring the identifiers, and the investigator may still need to provide complete answers for the data security questions in the protocol.

***The Departments within the California Health and Human Services Agency (CHHSA) are: Aging, Alcohol and Drug Programs, Child Support Services, Community Services and Development, Developmental Services, Emergency Medical Services Authority, Health Care Services, Mental Health, Public Health, Rehabilitation, Social Services and Statewide Health Planning and Development.*

Not applicable

ANTICIPATED PROJECT START DATE

Projects cannot begin before they have been reviewed. The earliest possible start date is always the date of the next public meeting at which the project will be heard.

Dear Researchers: This date is related to the start date of the initial submission and you do not need to change it when submitting an amendment. Thanks,

07/08/2024 • Sussan Atifeh • *Not* Internal

For a list of public meeting dates, see the CPHS website

02/14/2022

ANTICIPATED PROJECT END DATE

08/14/2023

Project Details

PURPOSE

Include a brief statement, less than 500 words, describing the research project. Be sure to address the background for the project, including relevant literature, the major research questions to be addressed, and the expected end product (e.g., article, report or other publications). Include the location(s) where the project will take place. The summary should be understandable to the general public.

Starting in this section and throughout the body of the application, when you describe changes, would you please restore the previous language (approved in the last amendment on 12/28/2023) unchanged and describe the changes at the end of the original text introduced with the note " June 2024 amendment update:". In this way the reviewer will be able to understand what you plan to do differently.

06/18/2024 • Sussan Atifeh • *Not* Internal

Please make clear what specific changes are being made in this amendment. If there are no changes related to the amendment please state "the amendment does not change the original statement of purpose".

07/17/2024 • Laura Lund, MA • *Not* Internal

Black/African American (AA) individuals develop non-small cell lung cancer (NSCLC) 5 years earlier than their European American (EA) counterparts. AAs also have a higher lung cancer incidence and mortality rates. The role of systemic racism in the development and maintenance of these disparities has been understudied. The objective of this proposal is to determine the impact of racism-related socio-environmental factors, including air pollution, traffic proximity, neighborhood deprivation, redlining and perceived discrimination, on the mutational signatures and the mutagenic process in the NSCLC tumors of AA patients. The central hypothesis is that measurable systemic racism stressors cause identifiable mutational signatures that impact disease progression in AA patients with early-stage NSCLC. Our study rationale is that genomic data from AA NSCLC specimens is lacking, which is hindering our understanding of the disease etiology and progression in this vulnerable population. For example, of over 1,100 NSCLC tumors in The Cancer Genome Atlas, only 54 are from AA patients. This work has the potential to identify specific pathways linking systemic racism stressors to worse outcomes, yielding targets for intervention and biomarkers for risk stratification. In this pilot study, we propose to survey 100 AA patients with stage I–II NSCLC from the Greater CA and LA Cancer Registry to determine their exposure to systemic racism stressors, which we will correlate with data from the whole-genome sequencing (WGS) of their tumor tissue performed at a third party WGS facility. WGS facility will only have de-

identified data. We will also perform WGS on 20 non-Hispanic white NSCLC patients as a comparison group. Our aims are to:(1) determine the extent to which exposure to structural racism over time is linked with differences in NSCLC tumor evolution by characterizing the types of mutations, the order of their acquisition, and the activity of mutational processes;(2) define the effect of structural racism stressors on early recurrence in stage I-II AA NSCLC patients; (3) determine the distribution of the genomics of never-smoking AAs with NSCLC compared to the Sherlock-Lung study&stressors of interest. Journal articles will be published on the results of the study. We will also explore whether differences between the WGS in AA and NHW NSCLC patients exist and whether any race-based variability may be driven by the differential exposure to the aforementioned stressors.

MAJOR RESEARCH QUESTION

What is the major research question to be addressed in this project?

Our aims are to: (1) identify and correlate systemic racism stressors faced by AA patients with NSCLC in our cohort; (2) determine the extent to which exposure to systemic racism stressors is associated with differences in the order and type of mutations and the mutational processes active during NSCLC tumor evolution; and (3) evaluate the associations between the mutation signatures, systemic racism streshsors, and early NSCLC recurrence.

July 2024 Amendment Update:
Our aims are to: (1) determine the extent to which exposure to structural racism over time is linked with differences in NSCLC tumor evolution by characterizing the types of mutations, the order of their acquisition, and the activity of mutational processes; (2) define the effect of structural racism stressors on early recurrence (within 2 years of surgery) in stage I-II AA NSCLC patients and (3) determine the distribution of the genomics of never smoking AAs with NSCLC compared to the Sherlock-Lung study and if this distribution is affected by the stressors of interest.

STUDY PROCEDURES

Describe in detail all procedures for this research. Do not attach grant applications or similar documents. Information in this application must be sufficient to fully explain the procedures without such documents

You state: Upon receiving information from the cancer registry, the COH study team will screen eligible patients and send an introductory packet to potential participants.

Will all potential participants be mailed hard copy documents?

07/17/2024 • Laura Lund, MA • *Not* Internal

You state: Participants who contact study team with interest will be given information about the entire study from the study team.

Will this information be given orally over the phone? In writing or by email? Is there a script that will be used to ensure that all potential participants get the same information?

07/17/2024 • Laura Lund, MA • *Not* Internal

You state: If the participant has not received the introductory packet and expresses interest in participating, they will be sent an informed consent form sheet and HIPAA authorization for retrospective tissue sampling, the survey with cover letter, and a self-addressed stamped envelope.

--All participants should receive the same information that was in the original mailing, including the CCR brochure.

--Under what circumstances would potential participants contact you if they haven't been made aware of the study through the mailing packet?

07/17/2024 • Laura Lund, MA • *Not* Internal

You state: The consent form can also be signed using Docusign.

All the contact/recruitment procedures described in the amendment appear to refer to hard copy mailings and forms.

When would there be an opportunity to sign the consent form using Docusign? If there is an electronic version of the documents being sent to or received from participants please describe this in the procedures.

07/17/2024 • Laura Lund, MA • *Not* Internal

You state: If we do not hear from individuals within 2 weeks after sending the pre-contact letter...

You have not described a pre-contact letter. The amendment only describes the mailing packet with introductory letter, consent form, HIPAA authorization, etc. Is there a pre-contact letter prior to the mailing packet?

07/17/2024 • Laura Lund, MA • *Not* Internal

You state: In this study, we will focus on address at diagnosis and control for length of residency at that address in models. In subsequent studies, we will incorporate their address history to explore the timelines of mutational development in conjunction with historical environmental stressors related to systematic racism.

Are you seeking approval for the use of the collected address data for future studies?

07/17/2024 • Laura Lund, MA • *Not* Internal

The website outreach aspect of the amendment cannot be approved until the promotional materials/website content is available for review.

07/17/2024 • Laura Lund, MA • *Not* Internal

The first sentence of the contact letter states: Thank you for your interest in participating in our lung cancer research study.

However, it appears from the recruitment procedures you describe that the mailing packet will be sent unsolicited to individuals who have never heard of your study, based on cancer registry information. This will be the first time they have heard of your study. This language may be confusing for potential participants. Please revise.

07/17/2024 • Laura Lund, MA • *Not* Internal

Notes on the letter of introduction--

--Please tell individuals at the beginning of the letter how you got their information, provide them with a brief description of your study and why they are being asked to participate, and describe for them briefly what will be expected of them if they choose to participate.

--In the second sentence, please replace the language 'it is advised' to language that more clearly states 'if you wish to participate in this study, please review and sign the consent form before answering the survey questions', or language to that effect.

--There is nothing in this letter that tells them what to do with the consent form after they sign it.

--The letter should clearly describe all the materials that you have enclosed in the packet.

--There is no description of the HIPAA authorization form as such, no request for them to sign it, and no instructions for how to return it.

--By 'list of questions' are you referring to the survey questionnaire? Please make this clear to the participant.

--There should be an option in the letter for potential participants to opt out if they do not want to be contacted any further about the study with a telephone number and/or email and the name of the person to contact.

--I have left comments later in the protocol regarding the multiple introduction letters that have been submitted. Please revise and consolidate into a single introductory letter that clearly provides all the necessary information to the potential participant.

07/17/2024 • Laura Lund, MA • *Not* Internal

In the cover letter you tell participants that they can complete the survey on the computer or over the phone. If they choose one of these options how will you ensure that you obtain

informed consent prior to administration of the survey? Please describe your plans for implementing these alternative modalities in your procedures.

07/17/2024 • Laura Lund, MA • *Not* Internal

You are relying on individuals to follow instructions and return a signed consent form with the survey. How will you handle situations in which an individual returns a survey without a signed consent form?

07/17/2024 • Laura Lund, MA • *Not* Internal

You are requesting to allow individuals to sign the informed consent without any consultation with project staff unless they reach out for assistance. This study involves collecting identifiable potentially sensitive information. In this case, at a minimum, your cover letter should clearly explain what informed consent is, what information is in the consent form, and why it is important that they should read it carefully and consult with project team members if they have any questions before signing.

07/17/2024 • Laura Lund, MA • *Not* Internal

You state: After receipt of the completed survey, HIPAA authorization, and consent documents, we will send the participants a $75 gift card ($25 for the questionnaire, $50 for the signed informed consent and HIPAA Authorization).

Why is the $75 broken down in this way? Will you provide $25 if they return the questionnaire but do not return the consent form or HIPAA authorization? Will you provide $50 if they return informed consent and HIPAA authorization but no questionnaire? What if they return the survey and the signed consent form but no HIPAA authorization?

07/17/2024 • Laura Lund, MA • *Not* Internal

Throughout the study, we will obtain names, contact information, sex, age, disease stage, and NSCLC surgical procedure from the Los Angeles County Registry. Upon receiving information from the cancer registries, the study team will contact and screen eligible patients and obtain informed consent and HIPAA authorization. The informed consent and HIPAA forms will then be provided to the registry, who will reach out to the institution where the tumor tissue resides to request tumor and non-involved lung tissue blocks from the definitive NSCLC resection. If blocks cannot be sent, we will ask for 20 tumor and 10 normal lung tissue unstained formalin-fixed paraffin-

embedded (FFPE) slides at 5 µm.

Recruitment activities: The challenges associated with recruiting AAs to participate in cancer research are well-documented. To enhance participation, evidence-based, culturally informed strategies from previous research will be implemented using a multi-pronged recruitment approach.

Recruitment & Retention Plan: Briefly, the study team will send a pre-contact letter, with PI contact information, to inform all potential participants about the study and offer them the opportunity to contact the study team if they wish to participate. Additionally, brochures that explain whole genome sequencing (WGS) and the California Cancer Registry (CCR) will be included.

Participants who contact study team with interest will be given information about the entire 2-part study from the study team. After participants have expressed interest in participating, they will be sent a consent information sheet for the survey that includes the waiver of documentation language, the survey with cover letter and a self-addressed stamped envelope. Once they receive these documents, a phone appointment will be set up to review the consent information sheet with the participant and answer any questions. Participants can complete the survey on line, over the phone or via paper. Additionally, they will be informed that they will be contacted for the second part of the study.

If we do not hear from individuals within 2 weeks after sending the pre-contact letter, we will call them no more than 3 times asking if they want to participate. Those who we contact that are interested, will be led through the same process as the above. Those who are not interested, will no longer be contacted.

After receipt of the completed survey, we will send the participants a $25 gift card. In addition, these individuals will be sent a medical release form, HIPAA authorization form and an informed consent form. Once they receive these documents (either via mail or electronically), study team will contact the participant to explain the forms and process and a phone/virtual appointment will be set up to review the consent form with the participant and answer any questions. Participants who wish to continue will be requested to provide signed consent to obtain tissue sampling. A self-addressed stamped envelope will be sent with the informed consent, HIPAA authorization and medical release forms but the consent form can be signed using Docusign, e-consent, or sent through the mail using a postage-paid envelope. After receiving these signed documents, we will send another $25 gift card.

Sociodemographic and clinical data: We will obtain individual-level data on demographic, clinical, and disease characteristics (including pathology reports) for each individual from the cancer registry.

Survey administration: Participants will complete an IRB-approved survey, developed from validated, high-quality standard measures of SDOH in the PhenX Toolkit. As noted above, participants can complete the survey either

online, via phone, or by mail. Our tool will obtain sociodemographic factors, including race, age, sex, education, annual household income, occupation, and insurance type. It will also include questions on tobacco use (e.g., pack year, product type, quit time), social determinants (e.g., housing, financial, and food insecurity), and perceived discrimination using the 9-item Major Discrimination Scale.

Confirmation of residential history: Our survey will ask participants to provide their addresses for the 20 years prior to their diagnosis and the length of time they have spent living at their residence at time of diagnosis. In this study, we will focus on address at diagnosis and control for length of residency at that address in models. In subsequent studies, we will incorporate their address history to explore the timelines of mutational development in conjunction with historical environmental stressors related to systematic racism.

Full patient address data is the ideal format for geocoding patients to their correct location, and thus obtaining the most accurate spatial data associated with their immediate surroundings. Several measures we plan on using are calculated at the census block group level. This is a very small spatial unit (containing between 600 ad 3000 people) where values can vary drastically from block group to block group, esp when considering measures of segregation and deprivation. Geocoding to street name without number can be problematic in the LA region where many streets stretch for miles. To ensure patients are protected, we will assign all patients a unique identifier. Address data will be separated from all other patient data and will only be associated with the unique identifier. Address data will be stored separately under password protection. All geocoding and geospatial analysis will be conducted on the address + unique identifier data only, with no matching patient information. Once spatial attributes are extracted (e.g., aggregate pollution level for a neighborhood or census block group deprivation score), the spatial attributes will be rejoined to the patient dataset without address information.

Geospatial analysis of pollution/traffic exposure and neighborhood deprivation: The residential addresses of the patients will be geocoded using HIPAA-compliant spatial analytics software (Spatialitics, Oak Brook, IL) and matched with data from the Environmental Justice Screening and Mapping Tool (EJSCREEN). This tool created by the EPA leverages data from the US Census, the EPA, and the US Department of Transportation to track environmental and health risks across the country. Environmental risk factors to be used in our analysis include PM2.5, NO2, and traffic proximity at the block group level for years 2010, 2015, 2020.

In addition, we will develop novel spatially and temporally refined measures of pollution and traffic exposure that will allow for a more precise assessment of the impact of environmental indicators on mutational signatures. For air pollution, we will develop an IDW model, as in our preliminary studies, to predict past pollutant concentrations at each patient's residence at diagnosis from 1990 to the present, depending on their length of residency. Furthermore, we will develop state-of-the-art land-use regression (LUR) models to interpolate residential exposure to PM2.5 and

NO2, based on modern machine learning algorithms and following approaches developed elsewhere using EPA AQS monitoring stations, distance to nearest highway, Aerosol Optical Depth (AOD) products from Multi-Angle Implementation of Atmospheric Correction (MAIAC), daily meteorological variables (temperature, humidity, and precipitation considering lagged exposures), calendar variables, and land use characteristics (e.g., population density, altitude, neighborhood income, racial/ethnic composition, rural/urban, and area deprivation indexes). We will develop our predictive LUR models using Distributed Random Forest and Gradient Boosting machine learning algorithms and will optimize their performance through cross-validation and by comparing predictive estimates to estimates observed at EPA AQS monitors. We will obtain daily 100x100m grid cells that we will use to assign residential exposure for each participant. EPA AQS monitoring estimates are available after just days, so we can update our air pollution estimates up to the end of the study.

We have also developed a traffic density model using data from Kalibrate TraffiMetrix for 1995, 2000, 2005, and 2015. We will expand this model to include data from 2020 and to cover all of California. Traffic counts are given as point data on the road network, and IDW will be used to extrapolate a unified traffic density surface for California at 50-m resolution. Participants will be matched to the closest year of data.

Nine neighborhood disadvantage indices will be assigned to participants based on the census tract derived from their home addresses. These indices, from a variety of state and national public agencies, include: Area Deprivation Index (University of Wisconsin), California Healthy Places Index (Public Health Alliance of Southern California), Labor Market Engagement Index (Housing and Urban Development [HUD]), Low Poverty Index (HUD), Low Transportation Cost Index (HUD), Regional Opportunity Index (UC Davis), School Proficiency Index (HUD), Social Vulnerability Index (CDC), and Walkability Score (EPA). These validated indices are freely available through online data portals. Participants will be matched to the closest year of data available.

We will also consider a range of novel environmental justice metrics that address more complex social/cultural issues, such as racial residential segregation, income inequality, and exposure inequality. Census data already validated and aggregated at the ZIP code level will be used to calculate measures of race/ethnic neighborhood segregation that describe neighborhood "typology" (i.e., level and type of segregation): (a) evenness – degree to which the proportion of a particular racial/ethnic group reflects the group's relative proportion in the entire area; (b) isolation – degree of potential contact between members of the same racial/ethnic group in a particular area; (c) concentration – population density of a particular racial/ethnic group within an area; (d) centralization – spatial proximity of racial/ethnic groups to the center of a metropolitan area for urban settings; and (e) clustering – degree to which racial/ethnic neighborhoods are adjacent to one another.

Tissue procurement: The University of Southern California Translational Pathology Core "Population-Based Tissue Arm" has a slide procurement

service to retrieve the tumor tissue of patients who underwent surgery at any LA County hospital and have signed consent. FFPE tissue blocks or unstained slides will be retrieved with corresponding pathology reports, de-identified, and sent to the COH Pathology Research Services Core (PRSC).

DNA extraction: Dr. Leonidas Arvanitis (COH pathologist) will H&E-stain and review the FFPE slides. For tumors of sufficient size and purity (>70% malignant cells), he will identify area(s) of malignant cells, which will be dissected from unstained sections. The pathology core will perform DNA extraction from the dissected tissues using the Qiagen nucleic acid extraction kit for FFPE (Qiagen Inc). Although RNA will not be used in this application, it will be stored for future work. Based on previous experience, we expect to obtain approximately 1 µg of DNA, as measured on a Qubit fluorometer. DNA with an A260-to-A280 ratio between 1.8 and 2.0, a minimum concentration of 12 ng/µl, and a total concentration of 100 ng will be used for further analysis. Matched genomic DNA will also be obtained from non-malignant lung tissue using the same process.

Whole-Genome Sequencing (WGS): WGS will be performed using contract sequencing from a third-party commercial vendor like Novogene. Extracted DNA samples will be used for library preparation. We have published WGS data from fresh frozen and FFPE tumors samples.74 For FFPE material, DNA must first have the cross-links reversed. End-repair and ligation with standard Illumina linkers produces libraries that need to be amplified prior to sequencing on Illumina instruments in 2x150bp mode; thus, there is some duplication in the libraries. Total data generated will be 60X tumor, 30X normal (180 gigabases/90 gigabases), which allows for high-quality mutation calling at a balanced price point and is similar depth to that in PCAWG.

Alignment: The sequences will be processed using a set of project-specific pipelines for alignment, variant calling, and quality control. Pooled sample reads will be de-multiplexed and sorted using Picard. Reads will be aligned to the hg38 reference sequence from the Human Genome Reference Consortium using BWA. Duplicate reads will be marked using Picard tools. The Genome Analysis Tool Kit (GATK) will be used for local realignment of reads around insertions or deletions (indels) and base quality score recalibration (BQSR).

Mutation calling: MuTect and SomaticIndelDetector77,78 will be used to identify somatic SNVs and short indels, respectively. Mutation annotation will be performed using Oncotator.

Copy number analysis: Copy number calls will be established by combining the output of six individual callers into a consensus using a multi-tier approach yielding a copy number profile and purity and ploidy values.

Genetic ancestry analysis: Genetic variants from the WGS data of the non-tumor tissue will be called using HaplotypeCaller and GenotypeGVCF tools in the GATK software. Variants with no missing genotypes will be selected and used to perform multidimensional scaling analysis using PLINK software. Individual genetic ancestry will be determined for each subject to control for

potential confounding due to population admixture structure. Genetic ancestry estimation will be performed using individuals from 1000 genomes as the reference populations: Yoruba in Ibadan, Nigeria (YRI); Han Chinese in Beijing, China (CHB; a proxy for Indigenous American ancestry); and Northern and Western Europeans (CEU). Individual ancestry will be estimated from the WGS data using the Bayesian Markov Chain-Monte Carlo (MCMC) method implemented in the program STRUCTURE version 2.1. STRUCTURE will be run under the admixture model using prior population information and independent allele frequencies. Additionally, we will compute principal components using EIGENSTRAT. Members of our investigative team have considerable expertise in accounting for population-stratification and leveraging genetic ancestry in population-based studies.

Classification of mutagenic processes active in each sample: We will assess the goodness-of-fit for the COSMIC SNV signatures, including additional SNV, indel, and DBS signatures identified in the data. If the signatures do not fit well for many samples, we will use the methods from Alexandrov et al. to jointly call signatures in TCGA mutation catalogs.

Statistical modeling: Our primary goal is to determine whether exposure to environmental mutagens (e.g., PM2.5 and NO2), as well as less well-studied systemic racism stressors like neighborhood deprivation and perceived discrimination, are associated with mutational signatures in NSCLC patients. Our null hypothesis is that there is no significant difference in the risk of mutations with increasing exposure to racism-related stressors. For each binary outcome and social-environmental factor of interest, we will use multilevel logistic regression (with random effects on the intercept at the census tract level of residence), adjusting for a priori identified confounders, including sex, ancestry, tobacco use, stage, BMI, and length of residence at address at diagnosis, to test if the risks of driver mutation prevalence are associated with changes in racism-related socio-environmental factors. We will estimate marginal risk ratios and risk differences by calculating average marginal effects, which represent the change in the mean of the outcome per unit change in exposure and covariates. We will first assess the role of each systemic racism stressor separately and then conduct mixture models with BKMR and quantile-based g-computation. An interaction term will be added to models between exposure to environmental mutagens and ancestry to test for multiplicative interaction in their association with mutational signatures. Correlations between exposures will be examined to reduce the number of comparisons made. Any highly correlated variables ($r > 0.6$) will be pared down to a single variable, selecting based on a priori evidence of biologic plausibility. Multiple comparisons will be conducted across various genes and types of mutations, so type I error will be adjusted using the Benjamini-Hochberg procedure.

Consensus subclonal architecture: Consensus subclonal architectures will be obtained by integrating the output of 11 subclonal reconstruction callers, as described in our Nature paper.

Timing of gains: Timing analysis will be also completed using the approaches in Gerstung et al. Briefly, the number of mutation-containing reads, the inferred copy number of the region, the purity of the tumor, and the clonality

of the mutation are used to estimate the variant allele frequency of a mutation. The timing of each gain is inferred based on the expected number of mutations, depending on the copy number configuration: gain of one allele (2+1), gain of both alleles (2+2), and gain of one allele and loss of the other (2+0).

Timing of driver mutations: According to the previous analysis, mutations were classified into four different timing stages: early and late clonal, clonal "unspecified," and subclonal using MutationTimeR. The ORs of early/late clonal and clonal/subclonal will be calculated for driver mutations against the distribution of all other mutations present in fragments with the same copy number composition in the samples with each particular driver. The background distribution of these ORs will be assessed with 1,000 bootstraps.

Integrative timing: For each pair of driver point mutations and recurrent CNAs, an ordering will be established (earlier, later, or unspecified). The information underlying this decision will be derived from the timing of each driver point mutation, as well as from the timing status of clonal and subclonal copy number segments. The orderings will be aggregated across all samples, and a sports statistics model will be employed to calculate the overall ranking of driver mutations.

Timing of mutational signatures: Once a set of signatures is established, we will estimate signature activity in the three different sets of mutations (early clonal, late clonal, subclonal), allowing us to quantify signature activity changes over time. This approach is similar to what we applied to PCAWG.

Calculate mutagenic ordering, mutation process activity, and mutation process activity over time: The mutation timing and signature analysis will be merged to create timing maps.

Evaluation of early recurrence: Consent will be obtained from participants to review medical records related to their NSCLC diagnosis and treatment outcomes. We have defined early recurrence as evidence of local, regional, or distant disease, as noted on the patient's surveillance scan or in their clinician's note, occurring within two years of their definitive NSCLC resection. We will obtain the medical records for the two years after resection (so from 2017-2020) and catalog this data in a REDCap database.

Association between systemic racism stressors and early recurrence: We will investigate the relationship between the binary outcome of recurrence with systemic racism stressors using logistic regression. Our null hypothesis is that there is no significant difference in the risk of recurrence with increasing exposure to racism-related stressors.

Association between oncogenic driver events, mutational signatures, and early recurrence: We will investigate the relationship between early recurrence and tumor mutational signatures and the timing of driver mutation events. Specifically, we will determine the extent to which the risk of early recurrence is associated with the frequency of SBS, DBS, indel, and driver mutation events. Our null hypothesis is that there is no relationship between the risk of early recurrence and the frequency of these genetic

events. We will use multiple logistic regression to assess these associations, adjusting for a priori identified confounders, including sex, ancestry, tobacco use, stage, BMI, and length of residence at address at diagnosis.

Comparison of the mutation spectra of our cohort to the mutational signatures of pertinent environmental toxins: Non-negative matrix factorization (NMF) will be used to deconvolute a M x N matrix of mutation catalogs into a M x K matrix of mutational processes and a K x N matrix of mutational exposures, where N is the number of tumors, M is the number of mutation types, and K is the number of estimated mutational processes. NMF will be run on MATLAB using SigProfiler47 and the nnmf function from the MATLAB Statistics Toolbox. We will use six mutation types with 16 different trinucleotide contexts for a total of 96 mutational states. The number of possible signatures will vary from 1 to 10, and signature stability will be assessed via sampling, as described. Cosine correlations will be used to determine similarity with signatures from Kucab et al.

Association between putative environmentally induced mutational signatures and exposure to the systemic racism stressors: We will evaluate the extent to which exposure to environmental mutagens (e.g., PM2.5 and NO2), neighborhood deprivation, and perceived discrimination is associated with having putative environmentally induced mutational signatures in NSCLC. Our null hypothesis is that there is no significant difference in the risk of having an environmentally induced mutational signature based on exposure to racism-related stressors.

Association between putative environmentally induced signature and early recurrence: We will evaluate the extent to which having putative environmentally induced mutational signatures is associated with early recurrence. Our null hypothesis that the risk of early recurrence is not associated with having the mutational signatures. We will then conduct a series of logistic regressions, evaluating the association between the signatures and the presence of early recurrence, adjusting for age, sex, stage, tobacco use, income, education, occupation, insurance type, financial and food security, ancestry, and exposure to PM2.5 and neighborhood deprivation. Multiple comparisons will be conducted across signature-specific regressions, so type I error will be adjusted for using the Benjamini-Hochberg procedure.

July 2024 Amendment Update:

Throughout the study, we will obtain names, contact information, sex, age, disease stage, and NSCLC surgical procedure from the Los Angeles County Registry and Cancer Registry of Greater CA. Upon receiving information from the cancer registries, the study team will screen eligible patients and obtain informed consent and HIPAA authorization. The completed informed consent and HIPAA forms will then be provided to the institution where the tumor tissue resides to request tumor and non-involved lung tissue blocks from the definitive NSCLC resection. If blocks cannot be sent, we will ask for 20 tumor and 10 normal lung tissue unstained formalin-fixed paraffin-embedded (FFPE) slides at 5 μm.

Recruitment activities: The challenges associated with recruiting AAs to participate in cancer research are well-documented. To enhance participation, evidence-based, culturally informed strategies from previous research will be implemented using a multi-pronged recruitment approach. We also will develop promotional materials, such as a website, to further engage our eligible population in our study. No PHI will be used in the promotion of this study, as the goal is to provide a general study overview that sparks genuine interest from prospective participants. We will not use any promotional materials prior to the recruitment materials being approved by CPHS.

Recruitment & Retention Plan: Upon receiving information from the cancer registry, the COH study team will screen eligible patients and send an introductory packet to all potential participants. The packet outlines details of the study and contains an introductory letter from the PI, study staff contact information in case participants have questions or would like to speak with a research staff member about the study, informed consent and HIPAA authorization forms for retrospective tissue sampling, the survey with cover letter, and a self-addressed stamped envelope. Once received, the COH staff will provide the completed consent documents for Greater CA Registry patients to PHI. Public Health Insititute (PHI) will provide the informed consent and HIPAA forms to the registries, who will reach out to the institution where the tumor tissue resides to request tissue. Potential participants have the option to contact the study team if they wish to participate. Additionally, brochures that explain whole genome sequencing (WGS) and the California Cancer Registry (CCR) will be included.

The consent form can also be signed using Docusign, which will be included in the introductory letter. Patients will be notified in the letter that if they wish to complete the forms via Docusign, they can contact COH study staff directly at the phone number provided in the letter. Once they receive these documents, participants have the option to set up a phone appointment with study staff to review the consent information sheet and answer any questions. Participants can complete the survey online, over the phone or via paper. Patients who return the completed survey, signed consent forms, and signed HIPAA authorization will be considered enrolled in the study. After receipt of the completed survey, HIPAA authorization, and consent documents, we will send the participants a $75 gift card ($25 for the questionnaire, $50 for the signed consent and HIPAA Authorization). Since the study is now compiled into "one part" to iinclude the questionnaire, informe onset, and HIPAA authorization, the reason for this is we lost 40% of patients to follow up from Part 1 to Part 2 (patients completed part 1 but did not compete part 2). For this reason, they are receiving a total of one $75 gift card to account for completing all study activities. Additionally, we have chosen to provide participatns with a generalized Visa giftcard, rather than a specialty (ie. Amazon) gift card due to our understanding patients have varying financial needs and the Visa gift card will be beneficial for patients in that they can use it at any establishment that accepts Visa cards. There is also additional costs due to this, since a Visa giftcard requires an activation coe and due to this additional cost we have consoldatied the gift card into one $75 gift card for this reason.

If we do not hear from individuals within 2 weeks after sending the introductory packet, we will call them no more than 10 times asking if they want to participate, in addition to resending the introductory packet. Those who we contact that are interested, will be led through the same process as the above. Those who are not interested, will no longer be contacted.

Sociodemographic and clinical data: We will obtain individual-level data on demographic, clinical, and disease characteristics (including pathology reports) for each individual from the cancer registries.

Survey administration: Participants will complete an IRB-approved survey, developed from validated, high-quality standard measures of SDOH in the PhenX Toolkit. As noted above, participants can complete the survey either online, via phone, or by mail. If they choose to complete the survey via phone or online, the only way to do this is to contact study staff. When patients reach out for the link or assistance with the survey over the phone, we will confirm with the patients that they have reviewed and signed the informed consent form. If they have not completed the informed consent form, they will be guided through the informed consent by study staff. If an individual returns the survey without the signed consent form, their survey information will not be inputted into our system. They survey will be returned to the patient and we will inform them that in order for us to accept the survey, they will have to return the reviewed and signed informed consent form. We will also give them our contact information in case they have any questions regarding this process. Our tool will obtain sociodemographic factors, including race, age, sex, education, annual household income, occupation, and insurance type. It will also include questions on tobacco use (e.g., pack year, product type, quit time), social determinants (e.g., housing, financial, and food insecurity), and perceived discrimination using the 9-item Major Discrimination Scale.

Confirmation of residential history: Our survey will ask participants to provide their addresses for the 20 years prior to their diagnosis and the length of time they have spent living at their residence at time of diagnosis. In this study, we will focus on address at diagnosis and control for length of residency at that address in models. In subsequent studies, we will incorporate their coded address history to explore the timelines of mutational development in conjunction with historical environmental stressors related to systematic racism. Patient addresss history will be de-identified and stored as a geospatial code that only COH study staff will have access to.

Full patient address data is the ideal format for geocoding patients to their correct location, and thus obtaining the most accurate spatial data associated with their immediate surroundings. If participants are unable to provide accurate address history, we will use LexisNexis to determine their previous residence for precise geospatial analyses. Several measures we plan on using are calculated at the census block group level. This is a very small spatial unit (containing between 600 ad 3000 people) where values can vary drastically from block group to block group, esp when considering measures of segregation and deprivation. Geocoding to street name without number can be problematic in the LA region where many streets stretch for

miles. To ensure patients are protected, we will assign all patients a unique identifier. Address data will be separated from all other patient data and will only be associated with the unique identifier. Address data will be stored separately under password protection. All geocoding and geospatial analysis will be conducted on the address + unique identifier data only, with no matching patient information. Once spatial attributes are extracted (e.g., aggregate pollution level for a neighborhood or census block group deprivation score), the spatial attributes will be rejoined to the patient dataset without address information.

Geospatial analysis of pollution/traffic exposure and neighborhood deprivation: The residential addresses of the patients will be geocoded using HIPAA-compliant spatial analytics software (Spatialitics, Oak Brook, IL) and matched with data from the Environmental Justice Screening and Mapping Tool (EJSCREEN). This tool created by the EPA leverages data from the US Census, the EPA, and the US Department of Transportation to track environmental and health risks across the country. Environmental risk factors to be used in our analysis include PM2.5, NO2, and traffic proximity at the block group level for years 2010, 2015, 2020.

In addition, we will develop novel spatially and temporally refined measures of pollution and traffic exposure that will allow for a more precise assessment of the impact of environmental indicators on mutational signatures. For air pollution, we will develop an IDW model, as in our preliminary studies, to predict past pollutant concentrations at each patient's residence at diagnosis from 1990 to the present, depending on their length of residency. Furthermore, we will develop state-of-the-art land-use regression (LUR) models to interpolate residential exposure to PM2.5 and NO2, based on modern machine learning algorithms and following approaches developed elsewhere using EPA AQS monitoring stations, distance to nearest highway, Aerosol Optical Depth (AOD) products from Multi-Angle Implementation of Atmospheric Correction (MAIAC), daily meteorological variables (temperature, humidity, and precipitation considering lagged exposures), calendar variables, and land use characteristics (e.g., population density, altitude, neighborhood income, racial/ethnic composition, rural/urban, and area deprivation indexes). We will develop our predictive LUR models using Distributed Random Forest and Gradient Boosting machine learning algorithms and will optimize their performance through cross-validation and by comparing predictive estimates to estimates observed at EPA AQS monitors. We will obtain daily 100x100m grid cells that we will use to assign residential exposure for each participant. EPA AQS monitoring estimates are available after just days, so we can update our air pollution estimates up to the end of the study.

We have also developed a traffic density model using data from Kalibrate TraffiMetrix for 1995, 2000, 2005, and 2015. We will expand this model to include data from 2020 and to cover all of California. Traffic counts are given as point data on the road network, and IDW will be used to extrapolate a unified traffic density surface for California at 50-m resolution. Participants will be matched to the closest year of data.

Nine neighborhood disadvantage indices will be assigned to participants

based on the census tract derived from their home addresses. These indices, from a variety of state and national public agencies, include: Area Deprivation Index (University of Wisconsin), California Healthy Places Index (Public Health Alliance of Southern California), Labor Market Engagement Index (Housing and Urban Development [HUD]), Low Poverty Index (HUD), Low Transportation Cost Index (HUD), Regional Opportunity Index (UC Davis), School Proficiency Index (HUD), Social Vulnerability Index (CDC), and Walkability Score (EPA). These validated indices are freely available through online data portals. Participants will be matched to the closest year of data available.

We will also consider a range of novel environmental justice metrics that address more complex social/cultural issues, such as racial residential segregation, income inequality, and exposure inequality. Census data already validated and aggregated at the ZIP code level will be used to calculate measures of race/ethnic neighborhood segregation that describe neighborhood "typology" (i.e., level and type of segregation): (a) evenness – degree to which the proportion of a particular racial/ethnic group reflects the group's relative proportion in the entire area; (b) isolation – degree of potential contact between members of the same racial/ethnic group in a particular area; (c) concentration – population density of a particular racial/ethnic group within an area; (d) centralization – spatial proximity of racial/ethnic groups to the center of a metropolitan area for urban settings; and (e) clustering – degree to which racial/ethnic neighborhoods are adjacent to one another.

Tissue procurement: The University of Southern California Translational Pathology Core "Population-Based Tissue Arm" has a slide procurement service to retrieve the tumor tissue of patients who underwent surgery at any LA County hospital and have signed consent. Similarly, the research unit at the Public Health Institute's Cancer Registry of Greater California (PHI/CRGC) program and the GA registry have biospecimen procurement services to obtain tumor and normal tissue of participants who underwent surgery from the facilities within their catchment areas and have signed consent. FFPE tissue blocks or unstained slides will be retrieved with corresponding pathology reports, de-identified, and sent to the COH Pathology Research Services Core (PRSC).

Nine neighborhood disadvantage indices will be assigned to participants based on the census tract derived from their home addresses. These indices, from a variety of state and national public agencies, include: Area Deprivation Index (University of Wisconsin), California Healthy Places Index (Public Health Alliance of Southern California), Labor Market Engagement Index (Housing and Urban Development [HUD]), Low Poverty Index (HUD), Low Transportation Cost Index (HUD), Regional Opportunity Index (UC Davis), School Proficiency Index (HUD), Social Vulnerability Index (CDC), and Walkability Score (EPA). These validated indices are freely available through online data portals. Participants will be matched to the closest year of data available.

We will also consider a range of novel environmental justice metrics that address more complex social/cultural issues, such as racial residential

segregation, income inequality, and exposure inequality. Census data already validated and aggregated at the ZIP code level will be used to calculate measures of race/ethnic neighborhood segregation that describe neighborhood "typology" (i.e., level and type of segregation): (a) evenness – degree to which the proportion of a particular racial/ethnic group reflects the group's relative proportion in the entire area; (b) isolation – degree of potential contact between members of the same racial/ethnic group in a particular area; (c) concentration – population density of a particular racial/ethnic group within an area; (d) centralization – spatial proximity of racial/ethnic groups to the center of a metropolitan area for urban settings; and (e) clustering – degree to which racial/ethnic neighborhoods are adjacent to one another.

Tissue procurement: The University of Southern California Translational Pathology Core "Population-Based Tissue Arm" has a slide procurement service to retrieve the tumor tissue of patients who underwent surgery at any LA County hospital and have signed consent. FFPE tissue blocks or unstained slides will be retrieved with corresponding pathology reports, de-identified, and sent to the COH Pathology Research Services Core (PRSC).

DNA extraction: Dr. Leonidas Arvanitis (COH pathologist) will H&E-stain and review the FFPE slides. For tumors of sufficient size and purity (>70% malignant cells), he will identify area(s) of malignant cells, which will be dissected from unstained sections. The pathology core will perform DNA extraction from the dissected tissues using the Qiagen nucleic acid extraction kit for FFPE (Qiagen Inc). Although RNA will not be used in this application, it will be stored for future work. Based on previous experience, we expect to obtain approximately 1 µg of DNA, as measured on a Qubit fluorometer. DNA with an A260-to-A280 ratio between 1.8 and 2.0, a minimum concentration of 12 ng/µl, and a total concentration of 100 ng will be used for further analysis. Matched genomic DNA will also be obtained from non-malignant lung tissue using the same process.

Whole-Genome Sequencing (WGS): WGS will be performed using contract sequencing from a third-party commercial vendor like Novogene. Extracted DNA samples will be used for library preparation. We have published WGS data from fresh frozen and FFPE tumors samples. For FFPE material, DNA must first have the cross-links reversed. End-repair and ligation with standard Illumina linkers produces libraries that need to be amplified prior to sequencing on Illumina instruments in 2x150bp mode; thus, there is some duplication in the libraries. Total data generated will be 60X tumor, 30X normal (180 gigabases/90 gigabases), which allows for high-quality mutation calling at a balanced price point and is similar depth to that in PCAWG.

Alignment: The sequences will be processed using a set of project-specific pipelines for alignment, variant calling, and quality control. Pooled sample reads will be de-multiplexed and sorted using Picard. Reads will be aligned to the hg38 reference sequence from the Human Genome Reference Consortium using BWA. Duplicate reads will be marked using Picard tools. The Genome Analysis Tool Kit (GATK) will be used for local realignment of reads around insertions or deletions (indels) and base quality score

recalibration (BQSR).

Mutation calling: MuTect and SomaticIndelDetector will be used to identify somatic SNVs and short indels, respectively. Mutation annotation will be performed using Oncotator.

Copy number analysis: Copy number calls will be established by combining the output of six individual callers into a consensus using a multi-tier approach yielding a copy number profile and purity and ploidy values.

Genetic ancestry analysis: Genetic variants from the WGS data of the non-tumor tissue will be called using HaplotypeCaller and GenotypeGVCF tools in the GATK software. Variants with no missing genotypes will be selected and used to perform multidimensional scaling analysis using PLINK software. Individual genetic ancestry will be determined for each subject to control for potential confounding due to population admixture structure. Genetic ancestry estimation will be performed using individuals from 1000 genomes as the reference populations: Yoruba in Ibadan, Nigeria (YRI); Han Chinese in Beijing, China (CHB; a proxy for Indigenous American ancestry); and Northern and Western Europeans (CEU). Individual ancestry will be estimated from the WGS data using the Bayesian Markov Chain-Monte Carlo (MCMC) method implemented in the program STRUCTURE version 2.1. STRUCTURE will be run under the admixture model using prior population information and independent allele frequencies. Additionally, we will compute principal components using EIGENSTRAT. Members of our investigative team have considerable expertise in accounting for population-stratification and leveraging genetic ancestry in population-based studies.

Classification of mutagenic processes active in each sample: We will assess the goodness-of-fit for the COSMIC SNV signatures, including additional SNV, indel, and DBS signatures identified in the data. If the signatures do not fit well for many samples, we will use the methods from Alexandrov et al. to jointly call signatures in TCGA mutation catalogs.

Statistical modeling: Our primary goal is to determine whether exposure to environmental mutagens (e.g., PM2.5 and NO2), as well as less well-studied systemic racism stressors like neighborhood deprivation and perceived discrimination, are associated with mutational signatures in NSCLC patients. Our null hypothesis is that there is no significant difference in the risk of mutations with increasing exposure to racism-related stressors. For each binary outcome and social-environmental factor of interest, we will use multilevel logistic regression (with random effects on the intercept at the census tract level of residence), adjusting for a priori identified confounders, including sex, ancestry, tobacco use, stage, BMI, and length of residence at address at diagnosis, to test if the risks of driver mutation prevalence are associated with changes in racism-related socio-environmental factors. We will estimate marginal risk ratios and risk differences by calculating average marginal effects, which represent the change in the mean of the outcome per unit change in exposure and covariates. We will first assess the role of each systemic racism stressor separately and then conduct mixture models with BKMR and quantile-based g-computation. An interaction term will be added to models between exposure to environmental mutagens and

ancestry to test for multiplicative interaction in their association with mutational signatures. Correlations between exposures will be examined to reduce the number of comparisons made. Any highly correlated variables ($r > 0.6$) will be pared down to a single variable, selecting based on a priori evidence of biologic plausibility. Multiple comparisons will be conducted across various genes and types of mutations, so type I error will be adjusted using the Benjamini-Hochberg procedure.

Consensus subclonal architecture: Consensus subclonal architectures will be obtained by integrating the output of 11 subclonal reconstruction callers, as described in our Nature paper.

Timing of gains: Timing analysis will be also completed using the approaches in Gerstung et al. Briefly, the number of mutation-containing reads, the inferred copy number of the region, the purity of the tumor, and the clonality of the mutation are used to estimate the variant allele frequency of a mutation. The timing of each gain is inferred based on the expected number of mutations, depending on the copy number configuration: gain of one allele (2+1), gain of both alleles (2+2), and gain of one allele and loss of the other (2+0).

Timing of driver mutations: According to the previous analysis, mutations were classified into four different timing stages: early and late clonal, clonal "unspecified," and subclonal using MutationTimeR. The ORs of early/late clonal and clonal/subclonal will be calculated for driver mutations against the distribution of all other mutations present in fragments with the same copy number composition in the samples with each particular driver. The background distribution of these ORs will be assessed with 1,000 bootstraps.

Integrative timing: For each pair of driver point mutations and recurrent CNAs, an ordering will be established (earlier, later, or unspecified). The information underlying this decision will be derived from the timing of each driver point mutation, as well as from the timing status of clonal and subclonal copy number segments. The orderings will be aggregated across all samples, and a sports statistics model will be employed to calculate the overall ranking of driver mutations.

Timing of mutational signatures: Once a set of signatures is established, we will estimate signature activity in the three different sets of mutations (early clonal, late clonal, subclonal), allowing us to quantify signature activity changes over time. This approach is similar to what we applied to PCAWG.

Calculate mutagenic ordering, mutation process activity, and mutation process activity over time: The mutation timing and signature analysis will be merged to create timing maps.

prEvaluation of early recurrence: Consent will be obtained from participants to review medical records related to their NSCLC diagnosis and treatment outcomes. We have defined early recurrence as evidence of local, regional, or distant disease, as noted on the patient's surveillance scan or in their clinician's note, occurring within two years of their definitive NSCLC resection. We will obtain the medical records for the two years after

resection (so from 2017-2027) and catalog this data in a REDCap database.

Association between systemic racism stressors and early recurrence: We will investigate the relationship between the binary outcome of recurrence with systemic racism stressors using logistic regression. Our null hypothesis is that there is no significant difference in the risk of recurrence with increasing exposure to racism-related stressors.

Association between oncogenic driver events, mutational signatures, and early recurrence: We will investigate the relationship between early recurrence and tumor mutational signatures and the timing of driver mutation events. Specifically, we will determine the extent to which the risk of early recurrence is associated with the frequency of SBS, DBS, indel, and driver mutation events. Our null hypothesis is that there is no relationship between the risk of early recurrence and the frequency of these genetic events. We will use multiple logistic regression to assess these associations, adjusting for a priori identified confounders, including sex, ancestry, tobacco use, stage, BMI, and length of residence at address at diagnosis.

Comparison of the mutation spectra of our cohort to the mutational signatures of pertinent environmental toxins: Non-negative matrix factorization (NMF) will be used to deconvolute a M x N matrix of mutation catalogs into a M x K matrix of mutational processes and a K x N matrix of mutational exposures, where N is the number of tumors, M is the number of mutation types, and K is the number of estimated mutational processes. NMF will be run on MATLAB using SigProfiler47 and the nnmf function from the MATLAB Statistics Toolbox. We will use six mutation types with 16 different trinucleotide contexts for a total of 96 mutational states. The number of possible signatures will vary from 1 to 10, and signature stability will be assessed via sampling, as described. Cosine correlations will be used to determine similarity with signatures from Kucab et al.

Association between putative environmentally induced mutational signatures and exposure to the systemic racism stressors: We will evaluate the extent to which exposure to environmental mutagens (e.g., PM2.5 and NO2), neighborhood deprivation, and perceived discrimination is associated with having putative environmentally induced mutational signatures in NSCLC. Our null hypothesis is that there is no significant difference in the risk of having an environmentally induced mutational signature based on exposure to racism-related stressors.

Association between putative environmentally induced signature and early recurrence: We will evaluate the extent to which having putative environmentally induced mutational signatures is associated with early recurrence. Our null hypothesis that the risk of early recurrence is not associated with having the mutational signatures. We will then conduct a series of logistic regressions, evaluating the association between the signatures and the presence of early recurrence, adjusting for age, sex, stage, tobacco use, income, education, occupation, insurance type, financial and food security, ancestry, and exposure to PM2.5 and neighborhood deprivation. Multiple comparisons will be conducted across signature-specific regressions, so type I error will be adjusted for using the Benjamini-

Hochberg procedure.

Please upload here any tables or charts related to your study procedures and any materials (such as surveys or interview questions) that will be presented to participants.

Survey with Coverletter - July 2024	Questionnaires
WGS Survey with Cover Letter - March 2022	Questionnaires

RECORDING

Will audio or video recording occur?

No

DECEPTION

Will deception be used in this study?

No

CALIFORNIA HEALTH AND HUMAN SERVICES AGENCY (CHHSA) DEPARTMENTS LIST

Indicate any of the following CHHSA department(s)' involvement in providing research staff, funding and/or patients from State mental hospitals for this project.

Not applicable

STATE DEPARTMENT DATA/SPECIMENS

Choose the department(s) from which you are requesting data and/or specimens and provide the formal name of the database or specimen registry. After you have selected the department from the drop down and entered the formal name of the database or specimen registry, click 'add' and repeat to add additional data and/or specimens if applicable.

Agency	Provide the formal name of the data base or specimen registry.
California Department of Public Health	California Cancer Registry Within the CCR registry, we are requesting access to some patient data from Los Angeles Cancer Registry and Cancer Registry of Greater California (see CCR letter of support).

Study Population

POPULATION DESCRIPTION

Provide a full description of how human subjects will be involved in the research. Address characteristics of subjects such as: age; sex; ethnicity; and number of participants. Include requested participant number.

Patient recruitment: We will recruit AA and non-Hispanic white (NHW) patients who have undergone definitive resection of their stage I–II NSCLC between 2015 and 2018 without neoadjuvant therapy.

Inclusion and exclusion criteria: Participants will be eligible if they are English-speaking, self-identify as Black/AA or NHW; underwent definitive R0 NSCLC resection for stage I–II NSCLC between 2015 and 2018 without neoadjuvant therapy, and have sufficient tissue for analyses after pathologic assessment. Patients will be excluded if they have a diagnosis of carcinoid tumors, sarcomas, or a non-thoracic malignancy (except non-melanoma skin cancers) within the last 5 years.

We aim to perform analysis of the tumor and surveys of 40 AA and 20 nHW individuals and thus will budget for recruitment of 50 AA and 25 nHW to achieve the goal.

July 2024 Amendment Update:

Patient recruitment: We will recruit AA patients who have undergone definitive resection of their stage I–II NSCLC between 2015 and 2025 without neoadjuvant therapy.

Inclusion and exclusion criteria: Participants will be eligible if they are English-speaking, self-identify as Black/AA; underwent definitive R0 NSCLC resection for stage I–II NSCLC between 2015 and 2025 without neoadjuvant therapy, and have sufficient tissue for analyses after pathologic assessment. Patients will be excluded if they have a diagnosis of carcinoid tumors, sarcomas, or a non-thoracic malignancy (except non-melanoma skin cancers) within the last 5 years.

We aim to perform analysis of the tumor and surveys of 100 AA individuals.

DATABASE DETAILS

List the database(s) to be used and the time period(s) being requested. This may include requests for future data that is not available at this time.

List the variables being requested, including a brief description of each variable.
Justify the need for each variable and for the quantity of data being requested.
You may also attach a list of variables on the next question.

Also address if participants will be involved in any other studies.

We will obtain variables from the California Cancer Registry (CCR) for the selected cases. We are aiming to understand whether socioenvironmental variables that are experienced by AA at higher rates due to historic and current racist practices are associated with NSCLC mutational signatures and timing. Individual-specific data are required. We need to adjust for other clinical and demographic factors that have also been requested. For enrolling participants, we will need address and phone number so that we can contact them. Participants will not necessarily be involved in any other studies.

If you have a list of variables with the details requested in the above question, attach that here. If you provided all details on the database in the question above, skip this question.

list of variables Ca Ca Registry .xlsx List of Variables

RATIONALE

What is the rationale for studying the requested group(s) of participants?

Black/African American (AA) individuals develop non-small cell lung cancer (NSCLC) 5 years earlier than their European American (EA) counterparts. AAs also have a higher lung cancer incidence and mortality rates. The role of systemic racism in the development and maintenance of these disparities has been understudied. The objective of this proposal is to determine the impact of racism-related socio-environmental factors, including air pollution, traffic proximity, neighborhood deprivation, redlining and perceived discrimination, on the mutational signatures and the mutagenic process in the NSCLC tumors of AA patients. The central hypothesis is that measurable systemic racism stressors cause identifiable mutational signatures that impact disease progression in AA patients with early-stage NSCLC. Our study rationale is that genomic data from AA NSCLC specimens is lacking, which is hindering our understanding of the disease etiology and progression in this vulnerable population. For example, of over 1,100 NSCLC tumors in The Cancer Genome Atlas, only 54 are from AA patients. We also aim to determine whether any differences between the WGS in AA and NHW NSCLC patients exist and whether any race-based variability may be driven by the differential exposure to the aforementioned stressors.

RECRUITMENT DETAILS

Describe how potential subjects will be identified for recruitment. Examples include: class rosters; group membership; individuals answering an advertisement; organization position titles (e.g., presidents, web designers, etc.). How will potential participants learn about the research and how will they be recruited (e.g., flyer, email, web posting, telephone, etc.)?

What does this mean: Participants will also be provided with a copy of the signed informed consent form in the perspective language.

07/17/2024 • Laura Lund, MA • *Not* Internal

Can you please clarify--there seem to be three different 'introductory' letters--there are documents here entitled 'Coverletter and form instruction July 2024' and 'introductory letter' and there is also a 'survey and coverletter' provided in the procedures section. Three introductory letters seems confusing, especially since all three provide some aspects of information essential for potential participants but none contains all the necessary information. Please consider consolidating into a single introductory letter to assist potential participants in understanding what your study is about and what they are expected to do.

07/17/2024 • Laura Lund, MA • *Not* Internal

Please provide the consent video guide referenced in the letter.

07/17/2024 • Laura Lund, MA • *Not* Internal

In the 'coverletter' document you state: 2. Please look at the bottom right corner of the page. Confirm that your name and date of birth are correct. If it is not correct, call us or check the box below and send us this sheet we will send you new forms,

Please make clear that they should use the self-addressed stamped envelope to return the form to you (so that they do not incur any mailing costs) and that they should not proceed to complete the survey until they receive and sign the new consent forms.

07/17/2024 • Laura Lund, MA • *Not* Internal

Why is there a box on the consent form for 'individual obtaining consent' if the consent forms are self-administered by the participant? 'Individual obtaining consent' implies that a member of the research team administered the consent to the participant.

07/17/2024 • Laura Lund, MA • *Not* Internal

Important to remember: subjects cannot be contacted before IRB approval.

Throughout the study, we will obtain names, contact information, sex, age, disease stage, and NSCLC surgical procedure from the Los Angeles County Registry. Upon receiving information from the cancer registries, the study team will contact and screen eligible patients and obtain informed consent and HIPAA authorization. The informed consent and HIPAA forms will then be provided to the registry, who will reach out to the institution where the tumor tissue resides to request tumor and non-involved lung tissue blocks from the definitive NSCLC resection. If blocks cannot be sent, we will ask for 20 tumor and 10 normal lung tissue unstained formalin-fixed paraffin-embedded (FFPE) slides at 5 μm.

Briefly, the study team will send a pre-contact letter, with PI contact information, to inform potential participants about the study and offer them the opportunity to contact the study team if they wish to participate. Additionally, brochures that explain whole genome sequencing (WGS) and the California Cancer Registry (CCR) will be included.

Participants who contact study team with interest will be given information about the entire 2-part study from the study team. After participants have expressed interest in participating, they will be sent a consent information sheet for the survey that includes the waiver of documentation language, the survey with cover letter and a self-addressed stamped envelope. Once they receive these documents, a phone appointment will be set up to review the consent information sheet with the participant and answer any questions. Participants can complete the survey on line, over the phone or via paper. Additionally, they will be informed that they will be contacted for the second part of the study.

If we do not hear from individuals within 2 weeks after sending the pre-contact letter, we will call them no more than 3 times asking if they want to participate. Those who we contact that are interested, will be led through the same process as the above. Those who are not interested, will no longer be contacted.

After receipt of the completed survey, we will send the participants a $25 gift card. In addition, these individuals will be sent a medical release form, HIPAA authorization form and an informed consent form. Once they receive these documents (either via mail or electronically), study team will contact the participant to explain the forms and process and a phone/virtual appointment will be set up to review the consent form with the participant

and answer any questions. Participants who wish to continue will be requested to provide signed consent to obtain tissue sampling. A self-addressed stamped envelope will be sent with the informed consent, HIPPA authorization and medical release forms but the consent form can be signed using Docusign, e-consent, or sent through the mail using a postage-paid envelope. After receiving these signed documents, we will send another $25 gift card.

July 2024 Amendment Update:

Throughout the study, we will obtain names, contact information, sex, age, disease stage, and NSCLC surgical procedure from the Los Angeles County Registry and Cancer Registry of Greater CA. Upon receiving information from the cancer registries, the study team will contact and screen eligible patients and obtain informed consent and HIPAA authorization. The informed consent and HIPAA forms will then be provided to the registry, who will reach out to the institution where the tumor tissue resides to request tumor and non-involved lung tissue blocks from the definitive NSCLC resection. If blocks cannot be sent, we will ask for 20 tumor and 10 normal lung tissue unstained formalin-fixed paraffin-embedded (FFPE) slides at 5 μm.

Upon receiving information from the cancer registry, the COH study team will screen eligible patients and send an introductory packet to potential participants. The packet outlines details of the study and contains an introductory letter from the PI (with a picture of the PI, an AA researcher), study staff contact information, informed consent and HIPAA authorization forms, the survey with cover letter, and a self-addressed stamped envelope.

Potential participants at both sites will be encouraged to contact the PI and study team to receive more information. A self-addressed stamped envelope will be included to return the paper survey, informed consent, and HIPAA authorization. Additionally, brochures that explain whole genome sequencing (WGS) and the California Cancer Registry (CCR) will be included.

Participants who contact study team with interest will be given information about the entire study from the study team. If the participant has not received the introductory packet and expresses interest in participating, they will be sent an informed consent and HIPAA authorization form for retrospective tissue sampling, the survey with cover letter, and a self-addressed stamped envelope. The consent form can also be signed using Docusign. Once they receive these documents, participants have the option to set up a phone appointment with study staff to review the consent information sheet and answer any questions. Participants can complete the survey online, over the phone or via paper. Patients who return the completed survey, signed consent forms, and signed HIPAA authorization will be considered enrolled in the study. After receipt of the completed survey, HIPAA authorization, and consent documents, we will send the participants a $75 gift card ($25 for the questionnaire, $50 for the signed informed consent and HIPAA Authorization). Participants will also be provided with a copy of the signed informed consent form in the perspective language.

If we do not hear from individuals within 2 weeks after sending the

introductory letter, we will call them no more than 10 times asking if they want to participate. Those who we contact that are interested, will be led through the same process as the above. Those who are not interested will no longer be contacted.

Attach copies of all recruitment materials.

Cancer-Research-in-California-BROCHURE-ENGLISH for WGS study Jan 2022.pdf	Recruitment Materials
Coverletter and Form Instruction - July 2024	Recruitment Materials
First Contact with Potential Study Participant - March 2022	Recruitment Materials
Introductory Letter - July 2024	Recruitment Materials
Video Consent Script - July 2024	Recruitment Materials
WGS patient brochure Jan 2022.docx	Recruitment Materials
WGS pre study intro letter March 2022	Recruitment Materials

SCREENING

Will subjects be screened prior to entry into the research?

No

COMPENSATION

Will subjects be compensated for participating in the study?

Yes

Compensation type
Gift card

Explain the amount and schedule of compensation that will be paid for participation in the study. Include provisions for prorating payment. The amount should not be coercive.
Upon receiving the completed survey, we will send a $25 gift card and request informed consent and HIPAA authorization/ medical release forms to obtain tissue samples. After receiving these, we will send another $25 gift card.

July 2024 Amendment Update:

Upon receiving the completed informed consent form, survey, and HIPAA authorization, we will send a $75 gift card to the enrolled participant ($25 for the questionnaire, $50 for the signed consent and HIPAA Authorization).

STUDY DURATION

Estimate the probable duration of the entire study. This estimate should include the total time each subject is to be involved and the duration of each data collection about the subject.

E.G., This is a two-year study. Participants will be interviewed three times per year; each interview will last approximately two hours. Total approximate time commitment for participants is 12 hours.

The duration of the pilot study is 1.5 years. The participants will participate for an estimated one hour each in order to consent, complete the study questionnaire and medical release forms.

July 2024 Amendment Update:
The duration of the pilot study is 1.5 years. The data from the pilot study (obtained from this study) was used to apply for a grant, which has been awarded as the R37. The project timeline has extended another 5 years, ending in 2029.

The participants will participate for an estimated one hour each in order to consent, complete the study questionnaire, and complete the medical release forms.

Risks and Benefits

RISK DESCRIPTION

Provide a description of possible risks to participants: physical, psychological, social, economic, loss of data security, and/or loss of confidentiality. Describe and justify whether the research is minimal risk or greater than minimal risk.

Overall risks are minimal because data will be collected from surveys and medical record data and kept confidential. All data sources will be linked by a study id and no personal identifiers will be kept in analytic files. The greatest potential risk is loss of confidentiality related to personal health information.

Research materials obtained from participants will primarily consist of self-reported surveys, and clinical data from patients' medical charts. All data are collected solely for the purpose of this study, and only the PI and key personnel will have access to private information pertaining to participants enrolled in the study.

All surveys are selected to minimize respondent burden. Perceived discrimination evaluations may involve some emotional distress as participants respond to questions. The risks from potential emotional distress from completion of the survey are expected to be minimal and are not a serious threat to the participants' health status.

Full patient address data is the ideal format for geocoding patients to their correct location, and thus obtaining the most accurate spatial data associated with their immediate surroundings. Several measures we plan on using are calculated at the census block group level. This is a very small spatial unit (containing between 600 ad 3000 people) where values can vary drastically from block group to block group, especially when considering measures of segregation and deprivation. Geocoding to street name without number can be problematic in the LA region where many streets stretch for miles. To ensure patients are protected, we will assign all patients a unique identifier. Address data will be separated from all other patient data and will only be associated with the unique identifier. Address data will be stored separately under password protection. All geocoding and geospatial analysis will be conducted on the address + unique identifier data only, with no matching patient information. Once spatial attributes are extracted (e.g., aggregate pollution level for a neighborhood or census block group deprivation score), the spatial attributes will be rejoined to the patient dataset without address information.

MEDICAL SERVICE RISKS

Describe how medical services will be provided if subjects suffer adverse mental or physical effects as result of research activity. If no services provided, state that clearly.

No medical services will be provided.

INTERNATIONAL RESEARCH

Will this research occur outside of the United States or U.S. territories?

Check with client to see if they consider territories to be outside the U.S. or not, as this can vary between institutions.

No

LESS RISKY METHODS

Describe any less risky methods and why they are not being used.

The least risky methods will be used in this project.

BENEFITS

Describe the benefits, if any, to the subjects or to society that will be realized as a result of this project. Discuss the benefits that may accrue directly to the subjects as well as to society. If there is no direct benefit anticipated for the subjects, state that clearly.

The results of this project will improve our understanding of the association between racism-related socioenvironmental factors and lung cancer mutation signatures in AA individuals. In particular, we may find targetable mutations that may help to identify a subset of AA at higher risk who may benefit from enhanced screening and or offered induction therapy.

Risks from survey completion are minimal, and participants are not expected to experience serious threats to health status by the completion of the surveys. The benefits outweigh the risks because risks are expected to be minimal, but the knowledge gained regarding the link between racism-related socio-environmental factors and lung cancer biology in AA individuals are expected to have far greater implications to public health.

There are no direct benefits anticipated for the participants.

JUSTIFICATION OF RISKS

Explain why study risks are reasonable in relation to the potential benefits to subjects and to society.

This is a low-risk study with the only potential risk being a loss of privacy; we have processes in place to prevent any disclosures of PHI. The study will likely not benefit participants directly but the results of this project will improve our understanding of the link between racism-related socio-environmental factors and lung cancer biology in AA individuals.

Adminstrative Safeguards

PERSONALLY IDENTIFIABLE DATA (PID) INSTRUCTIONS

Protected Health Information/Personally Identifiable Data (PHI/PID) is defined as information in any format that identifies the individual, including demographic information collected from an individual that can reasonably be used to identify the individual. Additionally, PHI is information created or received by a healthcare provider, health plan, employer, or health care clearinghouse; and relates to the past, present, or future physical or mental health or condition of an individual, including any of the 18 HIPAA identifiers.

Note: Please be aware that individual participants may be identifiable by combining other items in the data even when none of the 18 HIPAA identifiers are present. Thus, a study may still contain PID even after removing or never acquiring the identifiers, and the investigator may still need to provide complete answers for the data security questions in the protocol.

If the researcher demonstrates that he or she is unable to comply with any of the requirements below, he or she may request an exception from these requirements. The researcher should indicate any measures that will be taken to address this requirement. The exception request should be made in the text box of the corresponding requirement. An exception will only be granted if the researcher can demonstrate that adequate alternative measures have been taken to minimize risks so as to justify the exception.

HIPAA IDENTIFIERS

Please identify which HIPAA Identifiers you plan to request as part of your submission.

Name
Address (all geographic subdivisions smaller than state, including street address, city county, and zip code)
All elements (except years) of dates related to an individual (including birthdate, admission date, discharge date, date of death, and exact age if over 89)
Telephone numbers
Email address
Medical record number

TRAINING PROCEDURES

Describe the procedures for training all research staff who have access to PID on privacy and security. Indicate if staff are required to sign a confidentiality statement related to general use, security, and privacy.

A confidentiality pledge is signed upon employment, prior to being given access to any
confidential data. Research staff are also required to complete Collaborative Institutional
Training Initiative (CITI) Program training in Human Subjects Protection, Good Clinical Practice,
and Health Information Privacy and Security (HIPS) at the beginning of employment and then
every three years. Staff also are required to complete annual training on data privacy and
security. Staff are required to have IRB and HIPAA certification to participate in the study.

STAFF VETTING PROCEDURES

Describe procedures, either background check or thorough reference check, for vetting staff who will have access to PID.

Background and reference checks are conducted by Human Resources for candidates being
considered for employment.

SUPPORT LETTER

Obtain and submit a department support/data release letter.

This is a statement from the state agency or department you are receiving data from. It must be on that agency's/department's letterhead and should include both

1) *that the release of the desired data is legal and*

2) *that the entity is willing to release the desired data to you, the researcher. If you are not receiving data, this letter should indicate that you are supported.*

***For VSAC requests, if you do not have a Departmental Letter of Support (LOS)/Data Release, you may upload a copy of the Data Request Form (application) from the department to secure a review for the upcoming cycle. The protocol will not be approved until the LOS is uploaded to the protocol.*

Please also review the CPHS Statement for Birth and Death Data.

CPHS_LOS_Ehunmwunsee L	Department Letter of Support
SIGNED_CPHS_LOS_Erhunmwunsee L.pdf	Department Letter of Support

PREVENTING RE-USE AND UNAUTHORIZED ACCESS

Explain how you will ensure that data will not be reused or provided to any unauthorized person or entity.

Unauthorized means that the person or entity does not have a need to access the data for purposes of the research project approved by CPHS.

Electronic access to Protected Health Information requires passwords to be changed every 90
days. Data is stored in a HIPAA compliant building with keycard access control. Data exported
for analysis will not include personally identifying information, individuals are identified by a
coded number. When staff leave the employment of the City of Hope, electronic access is
removed and all keys/key cards are returned.

All data will be kept on secure servers and not released to any unauthorized person or entity.
Dr. Erhunmwunsee will assure that she will not release data for any other purpose by signing the
CCR's Appendix 3.

CONFIDENTIALITY OF PUBLISHED DATA

Indicate whether information will be published that could possibly be used to identify an individual subject.

No publications will include individual's names or allow for identification of an individual. All
results will be presented in tabular fashion.

DATA REQUEST JUSTIFICATION

Provide adequate justifications for the quantity of the data, the years and the variables being requested. Have you requested no more than the minimum necessary data to perform the research?

We are trying to understand the association of racism-related socio-environmental factors and AA early stage NSCLC mutational signatures and early recurrence. And we will explore whether there is a difference between WGS in NSCLC individuals based on race that is driven by racism-related socio-environmental stressors. It is important to be able to look more specifically at variables involved with diagnosis and treatment of lung cancer as there may be mutations/genes that are more associated with specific variables. Individual-specific data are required. In addition, specific mutations may be associated with outcomes of recurrence. Only needed variables will be requested.

LIMITATIONS TO DATA ACCESS

Indicate if access to data is limited only to those with a need to know for purposes of implementing or evaluating the research.

Access to data is limited to only those working on the project with a need to know for purposes
of implementing or evaluating the research.

PROTECTION AGAINST SMALL CELL SIZES AND ASSOCIATED PROBLEMS

Describe appropriate and sufficient methods to protect the identity of individual subjects when small cells or small numbers and/or data linkage to another data set are involved in the research project.

We will assess the potential risk that participants can be re-identified (e.g. Publication Scoring Criteria) and if high, combine categories and remove insignificant variables and apply other statistical masking that mitigates potential risk. Additionally, de-identification of data will be performed before public release of any data.

LINKAGES

Will the data set be linked with any other data sets?

Yes

Identify all data sets and each of the variables to be linked, with a brief description of each variable and justification for each linkage. If there is an extensive list, you may attach that list in the next question and indicate such here.
We will obtain vital statistics data linked to the cancer registry data.

Attach a copy of the document detailing all data sets and each of the variables to be linked. If you provided this information in the answer to the above question, skip this question.
No answer provided.

Will a third party be used for data linkage?
No

DESTRUCTION OF PID VERIFICATION

Indicate that you will provide CPHS with a letter certifying that PID has been destroyed and/or returned to the data source once research is concluded.

Yes

DATA SECURITY LETTER

If PHI will be handling the CCR data as a partner in this study please provide a data security letter for PHI. Please see my note earlier regarding PHI's role in this study.

07/17/2024 • Laura Lund, MA • *Not* Internal

Upload a certification/statement from the Chief Information Officer, Privacy Officer, Security Officer or equivalent position of the researcher's institution that CPHS Data Security Standards are met.

- *Data security letters cannot be signed by the Principal Investigator or Responsible Official.*
- *The data security letter must be on your institution's letterhead.*
- *Example of data security letter*

Data Security Requirement Letter LJE .docx Data Security Letter

Physical Safeguards

DATA PROTECTION

Indicate that research records and physical samples will be protected through the use of locked cabinets and locked rooms; PID in paper form will not be left unattended unless locked in a file cabinet, file room, desk, or office.

Yes

DATA DESTRUCTION

Will data/samples will be destroyed or returned as soon as it is no longer needed for the research project.

Yes

RETAINED DATA

Will the retained data/samples have personal identifiers or be de-identified?

data will contain personal identifiers

DESTRUCTION METHODS

Describe how you will ensure the PID in paper form is disposed of through confidential means, such as cross cut shredding or pulverizing.

All discarded PID are shredded using a professional shredding service.

FAXING

Describe how you will ensure that faxes with PID are not left unattended and fax machines are in secure areas.

Fax machines that are used for documents with PID are in secure areas.

MAILING

Indicate whether mailings of PID are sealed and secured from inappropriate viewing; and whether mailings of 500 or more individually identifiable records of PID in a single package, and all mailings of PID to vendors/contractors/co-researchers, are sent using a tracked mailing method, which includes verification of delivery and receipt, such as UPS, U.S. Express Mail, or Federal Express, or by bonded courier.

Any mailing of PID to request medical records would be sealed and protected, marked as
confidential, and sent with a tracking number. There would be no mailings of 500 or more of
individually identifiable records of PID.

ELECTRONIC STORAGE

State whether PID in paper or electronic form, e.g., stored on laptop computers and portable electronic storage media (e.g., USB drives and CDs), will ever be left unattended in cars or other unsecured locations.

Any PID on paper or electronic form stored on laptop computers or portable electronic storage media will never be left unattended to cars or other unsecured locations.

PHYSICAL STORAGE

Describe whether facilities, which store PID in paper or electronic form, have controlled access procedures, and 24 hour guard or monitored alarm service.

All facilities which store PID in paper or electronic form at COH are protected by controlled access procedures and have necessary protection as required.

SERVER SECURITY

Provide a description of whether all servers containing unencrypted PID are housed in a secure room with controlled access procedures.

All unencrypted PID are stored on servers located in secured rooms with controlled access procedures.
Files with PID which are not stored in secured rooms are encrypted.

STORING IDENTIFIERS

Indicate whether identifiers will be stored separately from analysis data.

Identifiers and analysis data will be stored together in a secure database with limited, password protected access. Only those recruiting patients can access identifying information. However, for analysis, the identifiers are not included in the data set.

The research database that the third party team will need for WGS analysis will have no personal identifiers included.

DISK STORAGE

State whether all disks with PID will be destroyed.

All disks with PID will be destroyed.

Electronic Safeguard

COMPUTER ACCESS OVERVIEW

State whether all computer access will be protected through the use of encryption, passwords, and other protections.

All computer access is protected through use of encryption, passwords and other protections.

FIPS 140-2 COMPLIANCE: WORKSTATIONS

Indicate whether all workstations that contain PID have full disc encryption that uses FIPS 140-2 compliant software. If not, explain why not and what encryption will be used.

All workstations that contain PID have full disc encryption using FIPS 140-2 compliant software.

FIPS 140-2 COMPLIANCE: LAPTOPS

Indicate if all laptops that contain PID have full disc encryption that uses FIPS 140-2 compliant software. If not, explain why not and what encryption will be used.

All laptops that contain PID have full disc encryption using FIPS 140-2 compliant software.

FIPS 140-2 COMPLIANCE: REMOVABLE MEDIA DEVICES

Indicate if PID on removable media devices (e.g. USB thumb drives, CD/DVD, smartphones, backup recordings) are encrypted with software that is FIPS 140-2 compliant.

Yes, all removable media devices are encrypted with software that is FIPS 140-2 compliant.

SECURITY PATCHES

Indicate if all workstations, laptops and other systems that process and/or store PID have security patches applied in a reasonable time frame.

All workstations, laptops and other systems that process and/or store PID have security patches applied promptly.

PASSWORD CONTROLS

Indicate if sufficiently strong password controls are in place to protect PID stored on workstations, laptops, servers, and removable media.

Yes sufficiently strong passwords are in place and are required to be changed on regular basis.
Passwords are required to be changed every 90 days. Our Password Policy requires a minimum of 8 characters with at least 3 of the following 4 requirements;
•One uppercase character
•One lowercase character
•One numeric character •One special character

ELECTRONIC SECURITY CONTROLS

Indicate if sufficient system security controls are in place for automatic screen timeout, automated audit trails, intrusion detection, anti-virus, and periodic system security/log reviews.

Electronic security controls are in place for automatic screen timeout, automated audit trails, intrusion detection, anti-virus, and periodic system security/log reviews.

FIPS 140-2 COMPLIANCE: ELECTRONIC TRANSMISSION

Explain whether all transmissions of electronic PID outside the secure internal network (e.g., emails, website access, and file transfer) are encrypted using software which is compliant with FIPS 140-2.

All transmission of electronic PID outside the secure internal network (e.g., emails, website access, and file transfer) are encrypted using software which is compliant with FIPS 140-2.

INTERNET ACCESSIBILITY

Note if PID in an electronic form will be accessible to the internet.

PID in an electronic form will not be accessible to the internet.

DISPOSING OF PID

When disposing of electronic PID, indicate whether sufficiently secure wiping, degaussing, or physical destruction will be used.

When disposing of electronic PID, sufficiently secure wiping, degaussing, or physical destruction will be used.

Conflict of Interest Information

CONFLICT OF INTEREST (COI) INSTRUCTIONS

A COI is defined as any financial or other relationships of the researcher(s) or the institution that could be perceived as affecting the objective conduct of the research, including the interpretation and publication of the findings. Researchers must disclose any COI, including perceived COI.

Financial relationships to be disclosed include but are not limited to the following:

- **Present or anticipated ownership of stock, stock options, or other financial obligations of the source of funding.**
- **Receipt or expectation of payment of any sort in connection with papers, symposia, consulting, editing, etc. from the source of funding.**
- **The sale or licensing or anticipated sale or licensing of medical or other products or intellectual property, such as patents, copyrights, or trade secrets to the source of funding or other entities.**
- **Any past, present or anticipated receipt of money or other valuable consideration from the source of research funding by the researcher(s), the family of the researcher(s), the research institution, or by an institution in which the researcher(s) or the family of the researcher(s) has an interest as owner, creditor, or officer.**

DISCLOSURES

Does any member of the study team, members' spouses, or members' dependent children have any significant financial interests related to the work to be conducted as part of the above-referenced project?

No

Informed Consent Procedures

INFORMED CONSENT PROCEDURES

Provide a description of procedures to be used in obtaining and documenting informed consent from participants.

> You state: Participants who contact the study team with interest will be given information about the entire study from the study team. We will describe the study, conduct informed consent, and offer the survey via phone, online through REDCap (Research Electronic Data Capture), or on paper.
>
> Elsewhere you have described consenting procedures that involve participants receiving a mailed packet, self-administering the consent form, and mailing it back to you. When you say here that you will conduct informed consent when they contact you, how will you do that?
>
> 07/17/2024 • Laura Lund, MA • *Not* Internal

See instructions and examples on CPHS website.

Consenting Plan: Upon receiving information from the cancer registry, the COH study team will screen eligible patients and send an introductory packet to potential participants. The packet outlines details of the study and contains an introductory letter from the PI, study staff contact information, informed consent and HIPAA authorization forms for retrospective tissue sampling, the survey with cover letter, and a self-addressed stamped envelope.

Potential participants have the option to contact the study team if they wish to participate. Additionally, brochures that explain whole genome sequencing (WGS) and the California Cancer Registry (CCR) will be included.

Participants who contact study team with interest will be given information about the entire study from the study team. If the participant has not received the introductory packet and expresses interest in participating, they will be sent an informed consent form sheet and HIPAA authorization for retrospective tissue sampling, the survey with cover letter, and a self-addressed stamped envelope. The consent form can also be signed using Docusign. Once they receive these documents, participants have the option to set up a phone appointment with study staff to review the consent information sheet and answer any questions. Participants can complete the survey online, over the phone or via paper. Patients who return the completed survey, signed consent forms, and signed HIPAA authorization will be considered enrolled in the study. After receipt of the completed survey, HIPAA authorization, and consent documents, we will send the participants a $75 gift card ($25 for the questionnaire, $50 for the signed consent and HIPAA Authorization).

If we do not hear from individuals within 2 weeks after sending the introductory letter, we will call them no more than 10 times asking if they

want to participate. Those who we contact that are interested, will be led through the same process as the above. Those who are not interested, will no longer be contacted.

July 2024 Amendment Update:

Consenting Plan: Upon receiving information from the cancer registries, the COH study team will screen eligible patients and send an introductory packet to potential participants. The packet outlines details of the study and contains an introductory letter from the PI, study staff contact information, informed consent and HIPAA authorization forms for retrospective tissue sampling, the survey with cover letter, and a self-addressed stamped envelope. Potential participants at both sites will be encouraged to contact the PI and study team to receive more information. Once they receive these documents, participants will review an IRB approved informed consent information form and study staff contact information.

Potential participants have the option to contact the study team if they wish to participate. Additionally, brochures that explain whole genome sequencing (WGS) and the California Cancer Registry (CCR) will be included.

Participants who contact the study team with interest will be given information about the entire study from the study team. We will describe the study, conduct informed consent, and offer the survey via phone, online through REDCap (Research Electronic Data Capture), or on paper. If the participant has not received the introductory packet and expresses interest in participating, they will be sent an informed consent form sheet and HIPAA authorization for retrospective tissue sampling, the survey with cover letter, and a self-addressed stamped envelope. The consent form can also be signed using Docusign. Once they receive these documents, participants have the option to set up a phone appointment with study staff to review the consent information sheet and answer any questions. Participants can complete the survey online, over the phone or via paper. Patients who return the completed survey, signed consent forms, and signed HIPAA authorization will be considered enrolled in the study. After receipt of the completed survey, HIPAA authorization, and consent documents, we will send the participants a $75 gift card ($25 for the questionnaire, $50 for the signed consent and HIPAA Authorization).

If we do not hear from individuals within 2 weeks after sending the introductory letter, we will call them no more than 10 times asking if they want to participate. Those who we contact that are interested, will be led through the same process as the above. Those who are not interested, will no longer be contacted.

CONSENT FORMS

Attach copies of consent forms and any other documents or oral scripts used to inform potential research subjects about the study. See examples of consent and assent forms on the CPHS website.

Please make the bill of rights a separate document (not part of the informed consent form) to be included in the mailing packet.

07/17/2024 • Laura Lund, MA • *Not* Internal

The consent form provided does not seem to match the instruction given for filling out the form. I did not see participant's name and DOB in the corner, and there is reference to a first page for COH use only that I could not find.

07/17/2024 • Laura Lund, MA • *Not* Internal

Using the readability score in Microsoft Word, the Flesch-Kincaid Grade Level for the consent form is 12.2. This is much too high for a self-administered consent form, especially given the sensitive and confidential nature of the medical and other data you are requesting. It is very likely that some participants will not understand the consent form completely. You may be able to reduce reading level to assist in potential participant comprehension if you eliminate passive sentences and reduce complex sentences by breaking them up into multiple simple sentences.

07/17/2024 • Laura Lund, MA • *Not* Internal

In the consent form, when you describe the risks associated with a breach of confidentiality, please describe for the participant what you are doing to mitigate those risks.

07/18/2024 • Laura Lund, MA • *Not* Internal

In the consent form, 'medically actionable' is a phrase that some participants may not understand. Please use lay terms.

07/18/2024 • Laura Lund, MA • *Not* Internal

Section VIII of the consent form refers to 'injury notification' but there is no language in the section describing injury notification procedures.

Also in the section it states that the PI '...has answered any and all questions'. I am not sure how this statement is possible since the participant has never heard of the study until opening the envelope and reading these materials.

07/18/2024 • Laura Lund, MA • *Not* Internal

The consent form states: Study participants will be compensated for their participation in this study. A $5750 gift card ($25 for completion of the survey and $50 for completion of the medical release forms) will be sent to the study participant.

Please see my earlier question regarding why the $75 gift is broken down this way.

07/18/2024 • Laura Lund, MA • *Not* Internal

Consent form--if participants withdraw their consent after initial participation, what will happen to the information you have already collected?

07/18/2024 • Laura Lund, MA • *Not* Internal

Consent form--under 'signature for consent', item 2 does not apply if the individual chose to self-administer the form and did not reach out to the project team to have the form explained. The instructional materials provided explain only how to fill out the form, not what the form means. Please remove or offer as an option only for those who were provided with assistance in understanding the form.

Item 4 states that they have been informed that they will receive a copy. When will this be done? I did not see this in the materials.

07/18/2024 • Laura Lund, MA • *Not* Internal

The HIPAA authorization should state specifically what information you will access from the medical record, for example, it should state that only information related to the diagnosis and treatment of the cancer will be released for the study.

07/18/2024 • Laura Lund, MA • *Not* Internal

The instructions given for filling out the HIPAA authorization

are different from the form presented here.

07/18/2024 • Laura Lund, MA • *Not* Internal

Be sure to include a concise explanation of key information for participants at the beginning of your consent form, as shown in the examples on the website. Also attach the Participant's Bill of Rights (download the revised version from the same CPHS website). CPHS may approve the use of a consent procedure which does not include, or which alters, some or all of the elements of informed consent. If a waiver or alteration of informed consent is being requested, attach a document that explains how all of the criteria below will be satisfied.

COH Bill Of Rights_Jan 2022.docx	Consent Form
COH HIPPA TEMPLATE Jan 2022.docx	Consent Form
Informed Consent - July 2024	Consent Form
March 2022 - COH Information Sheet - Survey Consent	Consent Form
March 2022 - Specimen Collection Adult - Minors WGS study consent	Consent Form
Medical Release Form WGS Jan 2022.doc	Consent Form
WGS consent form letter Jan 2022.docx	Consent Form

HIPAA Determination

HIPAA INSTRUCTIONS

To determine if this project is covered by HIPAA, answer the following questions.

COVERED ENTITY

Will health information be obtained from a covered entity, known as a clearinghouse, such as Blue Cross, that processes or facilitates processing health data from another entity, including but not limited to state databases?

No

HEALTHCARE PROVISIONS

Will the study involve the provision of healthcare by a covered entity, such as the UCD Medical Center?

No

OTHER HIPAA CRITERIA

Will the study involve other HIPAA criteria not listed above?

This study is covered by HIPAA as it accesses patient medical records.

07/18/2024 • Laura Lund, MA • *Not* Internal

No

Amendment Changes

List the pages and questions that have been changed.

Page 1 - Summary of Proposed Changes
Page 4 - Purpose, Major Research Quesion, Study Procedures, State Department Data/Specimens
Page 5 - Population Description, Recruitment Details, Study Duration
Page 7- Support letter, Data Security Letter
Page 11- Informed consent procedures, Consent Forms

Cover Letter and PI Signature for PI Submission

BUDGET

Does this project have a budget?

Please provide a budget for the amended version of this study. This budget appears to be for the pilot phase only, not the work being conducted for the amendment.

07/17/2024 • Laura Lund, MA • *Not* Internal

Yes

Attach a copy of your project budget here

CPHS budget Jan 2022.xls Project Budget

COVER LETTER

Attach a copy of your project cover letter.

Cover letter must have the requesting institution's letterhead.

Letter of Recommendation - Dr. Erhunmwunsee (California Health and Human Services Agency).pdf	Cover Letter

In order for the PI to review and sign this form, you will need to click "Next" and on the next page, click "Submit." At that point the PI will receive notification that will need to review the application and if they request changes, they will return the form to you and you will receive an email notification.

Calculated Field for agency plus data set *(Internal)*

California Department of Public Health: California Cancer Registry

Within the CCR registry, we are requesting access to some patient data from Los Angeles Cancer Registry and Cancer Registry of Greater California (see CCR letter of support).