ALLIANCE FOR CLINICAL TRIALS IN ONCOLOGY

ALLIANCE A151216

Adjuvant Lung Cancer Enrichment Marker Identification and Sequencing Trial (ALCHEMIST)
A screening trial for ALCHEMIST treatment trials
ClinicalTrials.gov Identifier: NCT02194738

Study Chair
Geoffrey Oxnard, MD
Lowe Center for Thoracic Oncology
Dana Farber Cancer Institute
450 Brookline Ave,
Boston, MA 02115
Tel: 617-632-6049
geoffrey_oxnard@dfci.harvard.edu

Surgical Co-chair
Dennis Wigle, MD
Mayo Clinic
Tel: 507-284-8462
wigle.dennis@mayo.edu

SWOG Co-Chair
David Gandara, M.D.
Tel: 916-734-5959
david.gandara@ucdmc.ucdavis.edu

ECOG Co-chair
Suresh Ramalingam
Tel: 404-778-7777
suresh.ramalingam@emory.edu

Disease Committee Chair
Thomas Stinchcombe, MD
Tel: 919-966-4432
thomas.stinchcombe@duke.edu

Primary Statistician
Sumithra Mandrekar, PhD
Tel: 507-266-6724
mandrekar.sumithra@mayo.edu

Secondary Statistician
Shauna Hillman, MS
Tel: 507-284-1533
hillman.shauna@mayo.edu

Protocol Coordinator
Colleen Watt
Tel: 773-702-4670 Fax: 312-345-0117
cboyle@uchicago.edu

Participating NCTN groups:
Alliance
ECOG-ACRIN
NRG
SWOG

Study Contacts:

Alliance Statistical and Data Center
Mayo Clinic
200 First St. SW
Rochester MN 55905

Protocol Resources:

A151216 Data Manager
Chris Bryhn
Tel: 507-284-0230
bryhn.christopher@mayo.edu

Laboratory Information

Cancer Genetics
Cancer Genetics, Inc
201 Route 17 North, 2nd Floor
Rutherford, NJ 07070
Tel: 201-528-9243
clinicaltrials@cgix.com

NCI Center for Cancer Genomics Biospecimen Core Resource
Julie M. Gastier-Foster, PI
Nationwide Children's Hospital
700 Children's Drive
Columbus, OH 43205
Tel: 614-355-3589
ALCHBCR@nationwidechildrens.org

CANCER TRIALS SUPPORT UNIT (CTSU) ADDRESS AND CONTACT INFORMATION

For regulatory requirements:	For patient enrollments:	For study data submission:
Regulatory documentation must be submitted to the CTSU via the Regulatory Submission Portal. (Sign in at www.ctsu.org and select the Regulatory > Regulatory Submission.) Institutions with patients waiting that are unable to use the Portal should alert the CTSU Regulatory Office immediately at 1-866-651-2878 to receive further instruction and support.	Refer to the patient enrollment section of the protocol for instructions on using the Oncology Patient Enrollment Network (OPEN) which can be accessed at: https://www.ctsu.org/OPEN_SYSTEM or https://OPEN.ctsu.org Contact the CTSU Help Desk with any OPEN-related questions at ctsucontact@westat.com	Data collection for this study will be done exclusively through Medidata Rave. Refer to the data submission section of the protocol for further information. Do not submit study study data or forms to the CTSU Data Operations. Do not copy the CTSU on data submissions.
The most current version of the **study protocol and all supporting documents** must be downloaded from the protocol-specific page located on the CTSU members' website (https://www.ctsu.org). Access to the CTSU members' website is managed through the Cancer Therapy and Evaluation Program - Identity and Access Management (CTEP-IAM) registration system and requires user log on with CTEP-IAM username and password. Permission to view and download this protocol and its supporting documents is restricted and is based on person and site roster assignment housed in the CTSU RSS.		
For clinical questions (i.e., patient eligibility or treatment-related) see the Protocol Contacts, Page 2.		
For non-clinical questions (i.e., unrelated to patient eligibility, treatment, or clinical data submission contact the CTSU Help Desk by phone or e-mail: CTSU General Information Line – 1-888-823-5923, or ctsucontact@westat.com. All calls and correspondence will be triaged to the appropriate CTSU representative.		
The CTSU Web site is located at https://www.ctsu.org.		

Patient Pre-Registration Eligibility Criteria
• **For pre-surgical patients**: Suspected resectable NSCLC and suspected stage large IB (≥ 4 cm), II (IIA or IIB) or IIIA using the 7th edition of the AJCC staging manual
• **For post-surgical patients:** Complete resection of NSCLC, path stage large IB (≥ 4 cm), II (IIA or IIB) or IIIA using the 7th edition of the AJCC staging manual
• **For all patients:**
   ECOG Performance Status 0-1; Age ≥18 years
   No patients who have received neoadjuvant therapy (chemotherapy or radiotherapy)
   No locally advanced or metastatic cancer requiring systemic therapy within 5 years; no secondary primary lung cancer concurrent or within 2 years.
   No prior treatment with agents targeting EGFR mutation, ALK rearrangement or PD-1/PD-L1/CTLA-4
   Non-pregnant and non-lactating
   Patients with local genotyping are eligible, regardless of the local result
   No recurrent lung cancer patients

Patient Registration Eligibility Criteria
• Adequate tissue available for the required analyses (either clinical tissue block or slides and scrolls)
• Completely resected stage large IB (≥4 cm), II (IIA or IIB) or IIIA using the 7th edition of the AJCC staging manual.
• **Non-squamous NSCLC (adjuvant therapy allowed)**.
• **Squamous cell NSCLC (no adjuvant therapy allowed)**
• Patients should be registered within specific timeframes, based on adjuvant chemo (see Section 3.2)
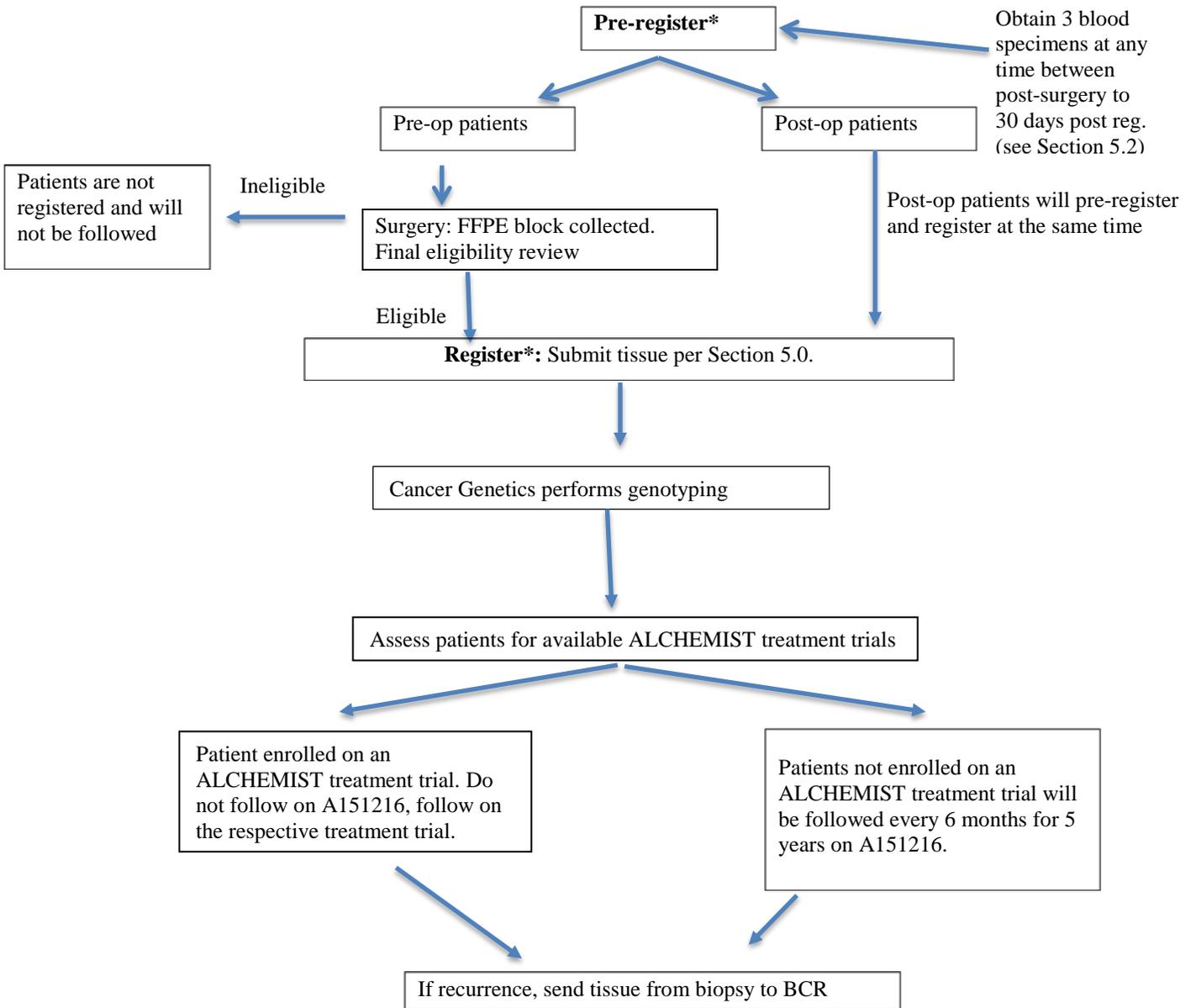
Schema

Pre-register*

Obtain 3 blood specimens at any time between post-surgery to 30 days post reg. (see Section 5.2)

Pre-op patients

Post-op patients

Patients are not registered and will not be followed

Ineligible

Surgery: FFPE block collected. Final eligibility review

Post-op patients will pre-register and register at the same time

Eligible

Register*: Submit tissue per Section 5.0.

Cancer Genetics performs genotyping

Assess patients for available ALCHEMIST treatment trials

Patient enrolled on an ALCHEMIST treatment trial. Do not follow on A151216, follow on the respective treatment trial.

Patients not enrolled on an ALCHEMIST treatment trial will be followed every 6 months for 5 years on A151216.

If recurrence, send tissue from biopsy to BCR

TABLE OF CONTENTS

1.0 INTRODUCTION

1.1 Selecting Therapies Based Upon Lung Cancer Genotype

The management of advanced lung adenocarcinoma has been transformed by the identification of targetable genotypes in a significant proportion of patients. Genotyping advanced lung adenocarcinomas for *EGFR* mutations and *ALK* rearrangements is now a routine part of care, as these genotypes indicate unique sensitivity to treatment with tyrosine kinase inhibitors (TKIs) such as erlotinib and crizotinib (1,2). However, these highly active targeted therapies do not lead to cures – resistance invariably develops.

1.2 Adjuvant Targeted Therapies

To determine whether highly active TKIs can improve cure rates, they must be studied in earlier stage disease (I-III). EGFR TKIs like erlotinib and gefitinib have been studied in the adjuvant setting, but not in genotype-selected populations (3,4). For this reason, studies randomizing genotype-selected lung cancer populations to TKI or placebo are under development both for erlotinib and crizotinib. To make these studies feasible, large numbers of resected lung cancers will need to undergo screening for *EGFR* mutations and *ALK* rearrangements. Because genotyping of resected lung cancers is not a routine part of clinical care, a screening trial is needed to identify *EGFR*-mutant and *ALK*-rearranged lung cancers.

1.3 Managing Resected Lung Cancer as a Genomically Diverse Disease

This study attempts to change the paradigm of adjuvant therapy in NSCLC towards the delivery of personalized genotype-directed therapies. The power of broad genomic characterization has already been demonstrated in advanced disease, highlighted by the work of the Lung Cancer Mutation Consortium (5). This collaboration of academic cancer centers committed to performing genotyping of 10 different genes for 1000 lung cancer patients, accelerating accrual to biomarker-based clinical trials. This current trial, A151216 (ALCHEMIST), now lays the groundwork for applying the same paradigm to resected lung cancer. The ALCHEMIST trial will screen patients with resected NSCLC using widely accepted genomic assays (*EGFR* mutation testing and *ALK* FISH) in a centralized CLIA-certified laboratory using resources from NCI. There are several advantages to this approach. First, centralized assays will be used for *EGFR* and *ALK* genotype analyses, minimizing technical inconsistencies. Second, the access to molecular testing will not be constrained by physician preferences or insurance approval. Third, additional genomic tests can be added to this platform over time to study other genotype-defined subtypes of NSCLC. Fourth, this trial will present an opportunity to characterize the natural history of NSCLC carrying less common genotypes (other than *EGFR*-mutant and *ALK*-rearranged NSCLC) through coordination with research genomics performed at Center for Cancer Genomics (CCG). Finally, working closely with CCG, the ALCHEMIST trial will facilitate large scale unbiased comprehensive molecular (using genome, exome and transcriptome) analyses to identify additional potentially targetable gene alterations.

1.4 Adjuvant Immunotherapy for Patients without a Targeted Treatment Option

Nivolumab is an FDA-approved drug for the 2nd line treatment of squamous cell carcinoma of the lung (6). It has also demonstrated improvement in overall survival when compared to docetaxel in the 2nd line treatment of non-squamous non-small cell lung cancer (7). In non-squamous non-small cell lung cancer, it is evident that selection of patients with tumors that express PD-L1, as detected on tumor cells by immunohistochemistry, can enrich for response to nivolumab. The same enrichment for clinical benefit has been demonstrated in all non-small-cell lung cancers (not selected by histology) treated with the anti-PD-1 agent pembrolizumab (8). There is not an absolute correlation between tumor PD-L1 expression and response to

nivolumab, but the difference in response rate and survival between cohorts of PD-L1 positive and PD-L1 negative non-squamous non-small cell lung cancer patients is significant.

1.5   Facilitating the Development of Next-generation Genomics

Next-generation genomics is increasingly being adopted into research and clinical efforts around the country, such that the limitations of DNA sequencing are slowly being discovered. To move beyond DNA sequencing and study gene expression and epigenetics, a rigorous central effort is needed. Already, a third generation of genomic technologies is in development that allows RNA sequencing and methylomics on paraffin embedded specimens. In collaboration with CCG, this study will be an opportunity to develop these technologies on clinically-annotated specimens, and to eventually explore the clinical significance of the results of this genomic research. Additionally, this study will create a unique opportunity to study change in genomics over time, through the collection and analysis of diagnostic specimens collected at recurrence.

1.6   ALCHEMIST Study Design

The ALCHEMIST study will accrue patients that are potentially eligible for the adjuvant treatment studies and perform molecular testing using a central reference laboratory certified by the Clinical Laboratory Improvement Amendments of 1988 (CLIA). Patients may either present prior to surgery with resectable NSCLC, or may present following complete resection (before or after adjuvant chemotherapy). All sites must submit patient's tissue for central genotyping. Patients will provide peripheral blood for matched normal DNA.

1.7   Central Clinical Genotyping

Non-squamous patients: All non-squamous patients (including those with local genotyping results) will have formalin-fixed tissue collected for central genotyping. The testing will be performed at Cancer Genetics, Inc., a commercial CLIA-certified laboratory. ALK FISH will be performed using the Vysis break-apart probe and *EGFR* genotyping will be performed by sequencing of exons 18-21.

Non-squamous and squamous patients will be tested for PD-L1 IHC staining at Cancer Genetics.

1.8   Advanced Genomics at CCG

In addition to the commercial genotyping at Cancer Genetics, tissue will be collected for research genomics by CCG. For those patients with a block available, this will be forwarded to the CCG after clinical testing at Cancer Genetics. For those patients without a block available, 10 micron scrolls (along with one H&E slide) should be cut from a block and submitted (the thicker sections reduce oxidative tissue damage seen with standard thickness slides). A peripheral blood specimen (EDTA) will be collected and sent to the CCG BCR to be used as a source of non-malignant ('germline') DNA. Also, two Streck tubes will be collected at baseline and shipped with the EDTA specimen. Please request blood kits from the CCG BCR through the BioMS system. Specimens will be coded. Over the course of the study, the CCG will perform advanced genomic analysis of the resected lung cancer specimens in a research, non-CLIA environment. Following completion of the genomic analysis, the results can be matched with the clinical follow-up results using a link between the samples coded and the patient identifiers for correlative analyses. The results of these genomic studies will not be provided back to the patient or their treating physician.

1.9   Recurrence Biopsy

Subjects participating in the follow-up portion of the ALCHEMIST study, as well as those participating in the adjuvant therapeutic studies, may, at the discretion of the treating physician,

undergo a standard-of-care diagnostic biopsy to confirm recurrence. If possible, core biopsies should be obtained as part of this recurrence biopsy. If available, paraffin embedded tissue from this biopsy should be sent to the NCI CCG BCR for additional research genomics (block preferred, scrolls with an H&E slide or unstained slides are acceptable). Please contact the BCR with any questions. In the event that re-biopsy tissue is not available, if clinical genomics testing is otherwise performed on the recurrence biopsy specimen, this data will be collected for research analysis as well.

2.0 OBJECTIVES

2.1 Primary Objectives

2.1.1 To centrally test resected NSCLC for genetic mutations to facilitate accrual to randomized adjuvant studies.

2.1.2 To obtain clinically annotated tumor tissue and patient-matched non-malignant DNA from peripheral blood, as well as detailed epidemiologic and clinical follow-up data, to allow clinically annotated advanced genomic analyses in concert with the NCI Center for Cancer Genomics (CCG).

2.2 Secondary Objectives

2.2.1 To characterize the natural history of molecularly characterized NSCLC to allow subsequent development of targeted therapies against genotype-defined subpopulations in the adjuvant and recurrent settings.

2.2.2 To cross-validate local genotyping assays for *EGFR* and *ALK* with a central reference standard.

2.3 Exploratory/Other Objectives

2.3.1 To study the genomic evolution of lung cancers by comparing genomic characteristics at resection and at recurrence.

2.3.2 To understand reasons behind lack of enrollment to adjuvant targeted therapy studies for potentially eligible patients.

2.3.3 To study the clinical significance of circulating tumor DNA within the plasma cell-free DNA (cfDNA) from early stage lung cancer patients.

3.0 PATIENT PRE-REGISTRATION/REGISTRATION ELIGIBILITY CRITERIA

3.1 Patient Pre-registration Eligibility Criteria

For pre-surgical patients

- Suspected diagnosis of resectable non-small cell lung cancer. Cancers with a histology of "adenosquamous" are considered a type of adenocarcinoma and thus a "nonsquamous" histology. Patients with squamous cell carcinoma are eligible.

- Suspected clinical stage of IIIA, II (IIA or IIB) or large IB (defined as size ≥4cm). Note: IB tumors <4cm are NOT eligible. Stage IB cancer based on pleural invasion is not eligible unless the tumor size is ≥4cm. **The 7th edition of AJCC staging will be utilized.**

For post-surgical patients

- Completely resected non-small cell lung cancer with negative margins (R0). Patients with squamous cell carcinoma are eligible only if they have not received adjuvant therapy.

- Pathologic stage IIIA, II (IIA or IIB) or large IB (defined as size ≥4 cm). Note: IB tumors <4cm are NOT eligible. Stage IB cancer based on pleural invasion is not eligible unless the tumor size is ≥4cm. **The 7th edition of AJCC staging will be utilized.**

For all patients

- ECOG Performance Status 0-1
- Age ≥ 18 years
- No patients who have received neoadjuvant therapy (chemo- or radio-therapy) for this lung cancer
- No locally advanced or metastatic cancer requiring systemic therapy within 5 years prior to registration. No secondary primary lung cancer diagnosed concurrently or within 2 year prior to registration.
- No prior treatment with agents targeting EGFR mutation, ALK rearrangement, and PD-1/PD-L1/CTLA-4.
- No patients known to be pregnant or lactating
- Patients who have had local genotyping are eligible, regardless of the local result.
- No patients with recurrence of lung cancer after prior resection.

Note: Post-surgical patients should proceed to registration immediately following pre-registration.

3.2 Patient Registration Eligibility Criteria

- Tissue available for the required analyses (either clinical tissue block or slides and scrolls, see Section 5.1)
- Completely resected NSCLC with negative margins (R0). Cancers with a histology of "adenosquamous" are considered a type of adenocarcinoma and thus a "nonsquamous" histology.
- Pathologic stage IIIA, IIA or IIB, or large IB (defined as size ≥ 4cm). Note: IB tumors <4cm are NOT eligible. Stage IB cancer based on pleural invasion is not eligible unless the tumor size is ≥4cm. **The 7th edition of AJCC staging will be utilized.**
- Patients with squamous cell carcinoma are eligible only if they have not received adjuvant therapy.
- In order to allow for time for central genotyping and eligibility for the ALCHEMIST treatment trial, patients must register within the following eligibility windows:

Squamous patients:

No adjuvant therapy permitted, register patient within 77 days following surgery

Non-squamous patients:

1. If no adjuvant therapy, register patient within 77 days following surgery.
2. If adjuvant chemotherapy **or** radiotherapy only, register patient within 225 days following surgery.
3. If adjuvant chemotherapy **and** radiation, register patient within 285 days following surgery.

4.0 PATIENT PRE-REGISTRATION AND REGISTRATION

Sites must have A151216 and the two treatment trials A081105 and E4512 IRB approved before registering patients to A151216.

All patients will pre-register to A151216. **Those patients that have already had surgery will complete the registration process at the same time as pre-registration**. Pre-op patients will be pre-registered and will then be registered following surgery, as long as all the registration eligibility criteria have been met. Non-squamous patients may be receiving adjuvant therapy at the time of registration to A151216.

4.1   CTEP Investigator Registration Procedures

Food and Drug Administration (FDA) regulations and National Cancer Institute (NCI) policy require all individuals contributing to NCI-sponsored trials to register and to renew their registration annually. To register, all individuals must obtain a Cancer Therapy Evaluation Program (CTEP) Identity and Access Management (IAM) account at https://ctepcore.nci.nih.gov/iam. In addition, persons with a registration type of Investigator (IVR), Non-Physician Investigator (NPIVR), or Associate Plus (AP) (i.e., clinical site staff requiring write access to OPEN, RAVE, or TRIAD or acting as a primary site contact) must complete their annual registration using CTEP's web-based Registration and Credential Repository (RCR) at https://ctepcore.nci.nih.gov/rcr.

RCR utilizes five person registration types.

- IVR — MD, DO, or international equivalent;

- NPIVR — advanced practice providers (e.g., NP or PA) or graduate level researchers (e.g., PhD);

- AP — clinical site staff (e.g., RN or CRA) with data entry access to CTSU applications (e.g., Roster Update Management System (RUMS), OPEN, Rave,);

- Associate (A) — other clinical site staff involved in the conduct of NCI-sponsored trials; and

- Associate Basic (AB) — individuals (e.g., pharmaceutical company employees) with limited access to NCI-supported systems.

RCR requires the following registration documents:

Documentation Required	IVR	NPIVR	AP	A	AB
FDA Form 1572	✔	✔			
Financial Disclosure Form	✔	✔	✔		
NCI Biosketch (education, training, employment, license, and certification)	✔	✔	✔		
HSP/GCP training	✔	✔	✔		
Agent Shipment Form (if applicable)	✔				
CV (optional)	✔	✔	✔		

An active CTEP-IAM user account and appropriate RCR registration is required to access all CTEP and CTSU (Cancer Trials Support Unit) websites and applications. In addition, IVRs and NPIVRs must list all clinical practice sites and IRBs covering their practice sites on the FDA Form 1572 in RCR to allow the following:

• Added to a site roster

• Assigned the treating, credit, consenting, or drug shipment (IVR only) tasks in OPEN

• Act as the site-protocol PI on the IRB approval

• Assigned the Clinical Investigator (CI) role on the Delegation of Tasks Log (DTL).

Additional information can be found on the CTEP website at https://ctep.cancer.gov/investigatorResources/default.htm. For questions, please contact the RCR Help Desk by email at RCRHelpDesk@nih.gov .

4.2 CTSU Site Registration Procedures

This study is supported by the NCI Cancer Trials Support Unit (CTSU).

IRB Approval:

For CTEP and Division of Cancer Prevention (DCP) studies open to the National Clinical Trials Network (NCTN) and NCI Community Oncology Research Program (NCORP) Research Bases after March 1, 2019, all U.S.-based sites must be members of the NCI Central Institutional Review Board (NCI CIRB). In addition, U.S.-based sites must accept the NCI CIRB review to activate new studies at the site after March 1, 2019. Local IRB review will continue to be accepted for studies that are not reviewed by the CIRB, or if the study was previously open at the site under the local IRB. International sites should continue to submit Research Ethics Board (REB) approval to the CTSU Regulatory Office following country-specific regulations.

Sites participating with the NCI CIRB must submit the Study Specific Worksheet for Local Context (SSW) to the CIRB using IRBManager to indicate their intent to open the study locally. The NCI CIRB's approval of the SSW is automatically communicated to the CTSU Regulatory Office, but sites are required to contact the CTSU Regulatory Office at CTSURegPref@ctsu.coccg.org to establish site preferences for applying NCI CIRB approvals across their Signatory Network. Site preferences can be set at the network or protocol level. Questions about establishing site preferences can be addressed to the CTSU Regulatory Office by emailing the email address above or calling 1-888-651-CTSU (2878).

In addition, the Site-Protocol Principal Investigator (PI) (i.e. the investigator on the IRB/REB approval) must meet the following criteria to complete processing of the IRB/REB approval record:

• Holds an Active CTEP status;

• Rostered at the site on the IRB/REB approval and on at least one participating roster;

• If using NCI CIRB, rostered on the NCI CIRB Signatory record;

• Includes the IRB number of the IRB providing approval in the Form FDA 1572 in the RCR profile; and

• Holds the appropriate CTEP registration type for the protocol.

4.2.1 Additional site registration requirements

Additional requirements to obtain an approved site registration status include:

• An active Federal Wide Assurance (FWA) number;

- An active roster affiliation with the Lead Protocol Organization (LPO) or a Participating Organization (PO); and

- Compliance with all protocol-specific requirements (PSRs).

4.2.2 Downloading Site Registration Documents

Download the site registration form from the A151216 protocol page located on the CTSU members' website. Permission to view and download this protocol and its supporting documents is restricted and is based on person and site roster assignment housed in the CTSU RSS. To participate, the institution and its associated investigators and staff must be associated with the LPO or a PO on the protocol.

- Log onto the CTSU members website (https://www.ctsu.org) and log in to the members' area using your CTEP-IAM username and password.

- Click on *Protocols* in the upper left of your screen.
    - Either the protocol number in the search field at the top of the protocol tree, or,
    - Click on the By Lead Organization folder to expand, then select trial protocol # A151216.

- Click on *Documents*, select *Site Registration*, and download and complete the forms provided.

4.2.3   Submitting Regulatory Documents

Submit required forms and documents to the CTSU Regulatory Office via the Regulatory Submission Portal on the CTSU website.

To access the Regulatory Submission Portal log onto the CTSU. members' website → Regulatory → Regulatory Submission

Institutions with patients waiting that are unable to use the Regulatory Submission Portal should alert the CTSU Regulatory Office immediately at 1-866-651-2878 in order to receive further instruction and support.

4.2.4 Checking Your Site's Registration Status

You can verify your site registration status on the members' section of the CTSU website.

Log on to the CTUS the members' website;

Click on *Regulatory* at the top of your screen;

Click on *Site Registration*;

Enter your 5-character CTEP Institution Code and click on Go.

Note: The status given only reflects institutional compliance with site registration requirements as outlined above. It does not reflect compliance with protocol requirements for individuals participating on the protocol or the enrolling investigator's status with the NCI or their affiliated networks.

4.3   Patient Registration Requirements

Informed consent: The patient must be aware of the neoplastic nature of his/her disease and willingly consent after being informed of the procedure to be followed, the experimental nature of the therapy, alternatives, potential benefits, side-effects, risks, and discomforts. Current human protection committee approval of this protocol and a consent form is required prior to patient consent and registration.

Patients with impaired decision-making capacity may be enrolled on this study, where institutional policy and IRB of record allow.

4.4   Study-specific Patient Registration Procedures

All patients will be pre-registered and registered to A151216 prior to submission of tissue for molecular testing. Patients can be pre-registered to A151216 pre-operatively or post-operatively.

Pre-operative patients:

Pre-operative patients will be pre-registered to A151216 prior to surgery. Following surgery and confirmation of registration eligibility criteria, patients will be registered to A151216 using the Alliance patient ID number assigned at pre-registration. This number will be entered into the OPEN registration system. You will then receive a confirmation of registration for your records.

Post-operative patients:

Post-op patients will pre-register and register to A151216 at the same time. Following confirmation of registration eligibility criteria, the CRA will register the patient by entering the Alliance patient ID number assigned at pre-registration into the OPEN registration system. The OPEN system will provide the institution with a printable confirmation of registration. Please print this confirmation for your records.

Registration Steps for Patients with Local Results

Patients with a local genotyping test results will pre-register and register to A151216 and may proceed to be screened for the appropriate treatment trial. All tissue and blood is still required for the genotyping and genomic research. These patients do not need to wait for confirmation of the positive test to enter the treatment trial. If the central laboratory (Central Genetics) does not confirm the local positive test, patients may still be registered to the appropriate treatment trial.

Epidemiological questionnaire for all patients: Following registration, a CRA will complete an epidemiological questionnaire to be used for comprehensive clinical annotation of the planned research genomics at CCG.

5.0   SPECIMEN COLLECTION AND SUBMISSION

5.1   Specimen Submission Overview and Timeline

Kits are available for the submission of whole blood (one EDTA and two Streck tubes included) to the BCR. Kits are not available for specimen submission to Cancer Genetics and the recurrence biopsy to BCR. **Kits for whole blood submission should be ordered using BioMS.** Each site can order up to 3 kits for collection. Kits should then be re-ordered once they have been used. Please note: Kits are available for whole blood collection only, not for tissue submission to Cancer Genetics or recurrence biopsy submission to the BCR.

Specimen	Timepoint	Ship to	Kit available?
1) Whole Blood (1 EDTA and 2 Streck tubes)	Obtain following surgery, up to 30 days following registration (30 days is preferred, but collection at a later time is acceptable), submit within 1 week of collection	BCR	Yes
2) Tissue Block	Submit following registration to A151216	Cancer Genetics	No
OR			

Unstained slides **AND** Scrolls (w/ one H&E slide)	Submit following registration to A151216	Unstained slides to Cancer Genetics **AND** Scrolls and H&E slide to BCR	No
3) Recurrence Biopsy (if done)	Submit at progression	BCR	No
A diagram of the submission process is included in Appendix I.			

1) **Blood:** All patients will have three whole blood specimens (one EDTA and two Streck tubes) obtained at any point after surgery up to 30 days following registration. Blood should be shipped to the BCR within 1 week of collection. It is preferred that blood be collected during the period between surgical resection and initiation of any adjuvant therapy. Later blood collection, including outside of the 30-day window following registration, is discouraged but will be accepted.

2) **Tumor blocks or cut slides and tissue scrolls:** Following registration, blocks or unstained slides will be submitted to Cancer Genetics laboratory for testing. If slides are submitted to Cancer Genetics then sites must submit tissue scrolls and an H&E slide to the BCR. (Tissue scrolls are "shavings" from the pathology block.)

3) **Recurrence Biopsy:** If a biopsy is done at recurrence tissue will be submitted to the BCR.

5.2 Blood Collection and Submission for All Patients

Blood should be obtained at any point after surgery up until 30 days following registration. It is preferred that blood be collected during the period between surgical resection and initiation of any adjuvant therapy. Later blood collection, including outside of the 30-day window following registration, is discouraged but will be accepted. Three whole blood specimens will be collected: one 10ml EDTA tube (provided in the kit) and two 10ml 10cc cell-free DNA blood collection tubes (BCT) from Streck (provided in the kit). **Please store all blood at room temperature prior to shipping.**

Ship at room temperature with the ambient packs included in the kits provided. Airbills may be requested for shipments to the NCI CCG by contacting: ALCHBCR@nationwidechildrens.org.

Please do not refrigerate or freeze the blood samples nor use dry ice for shipment.

NCI CCG Biospecimen Core Resource
Nationwide Children's Hospital
700 Children's Dr., WA1340
Columbus, OH 43205
Tel: 614-355-3589
ALCHBCR@nationwidechildrens.org

5.3 Tissue Preparation

See Appendix II for further information regarding block and slide submission to Cancer Genetics.

Study Tissue Block: At the time of surgical resection and gross pathology review, an additional segment of grossly apparent primary tumor tissue will be embedded for study purposes only. This tumor tissue block will not be returned to the site. Prior to distribution of the block, the site should cut, stain, and review one section of the study block to confirm that it is representative of the clinical diagnosis document tumor cellularity by histological

review. Cancer Genetics will then forward the remaining block to the BCR for genomic research.

OR

Clinical Tissue Block: The site pathologist should identify one block of primary tumor tissue from the case that is representative of the histological diagnosis, contains at least 1 cm^2 of tissue on the block face, and document tumor cellularity by histological review. Cancer Genetics will then forward the block to the BCR for genomic research. Note that if the block is required by the site at some future date for clinical patient management, the block will be returned to the site within upon a written request, if physically possible, but this cannot be guaranteed.

OR

Unstained Tissue Slides and Tissue Scrolls

Unstained Tissue Slides: The site pathologist should identify one block of primary tumor tissue from the case that is representative of the histological diagnosis and document tumor cellularity by histological review. A total of five (5) 10-micron sections plus eight (8) 5-micron sections should be cut and mounted on positively-charged glass slides and shipped to Cancer Genetics.

Tissue Scrolls (see scroll calculator on the Alliance or CTSU website to calculate amount of scrolls needed): The site pathologist should also identify one block of primary tumor tissue from the case that is representative of the histological diagnosis and document tumor cellularity by histological review (preferably, this will be the same block from which slides were cut for shipment to Cancer Genetics). One 5-6 micron section slide should be cut and stained with H/E, followed by a number of 10 micron tissue sections (scrolls, also known as "shavings") calculated as follows:

Number of tissue sections = $12 / (0.01 * L * W)$;

where L, W are the approximate cross-sectional length and width of the **tissue surface**, in mm.

Site pathologists may use the scroll calculator on the A151216 webpage to quickly determine the number of sections needed based upon tissue cross sectional area.

Scrolls should be sealed inside a microcentrifuge tube and a final 5-6 micron section should be cut and stained with H/E. The scroll and slide should be labeled with the A151216 patient ID and the protocol ID (A151216) prior to shipping. **The sealed tube of tissue scrolls and both H/E stained referenced slides should be shipped to the BCR on the same day of sectioning. Please contact the BCR for a priority overnight FedEx airbill.** To ensure rapid processing, note that blocks should not be sectioned or shipped on a Friday-Saturday, or a day before a holiday.

5.4   Specimen Submission using the Alliance Biospecimen Management System

USE OF THE ALLIANCE BIOSPECIMEN MANAGEMENT SYSTEM (BioMS) IS MANDATORY AND ALL SPECIMENS MUST BE LOGGED AND SHIPPED VIA THIS SYSTEM. **Additionally, sites will use BioMS to request kits for the blood draw.**

BioMS is a web-based system for logging and tracking all biospecimens collected on Alliance trials.   Authorized   individuals   may   access   BioMS   at   the   following   URL: http://bioms.allianceforclinicaltrialsinoncology.org using most standard web browsers (Safari, Firefox, Internet Explorer). For information on using the BioMS system, please refer to the 'Help' links on the BioMS web page to access the on-line user manual, FAQs, and training videos. To report technical problems, such as login issues or application errors, please contact:

1-855-55-BIOMS or Bioms@alliancenctn.org. For assistance in using the application or questions or problems related to specific specimen logging, please contact: 1-855-55-BIOMS or Bioms@alliancenctn.org.

After logging collected specimens in BioMS, the system will create a shipping manifest. This shipping manifest must be printed and placed in the shipment container with the specimens.

All submitted specimens must be labeled with the protocol number (A151216) and Alliance patient number.

The de-identified pathology report, being submitted with the sample to Cancer Genetics, should include the protocol number (A151216), patient number and patient's initials.

A copy of the Shipment Packing Slip produced by BioMS must be printed and placed in the shipment with the specimens.

Instructions for the collection of samples are included below. Please be sure to use a method of shipping that is secure and traceable. Extreme heat precautions should be taken when necessary.

Shipment on Monday through Thursday by overnight service to assure receipt is required. If shipping on Friday, FedEx or UPS must be used and the air bill must be marked "For Saturday delivery." Do not ship specimens on Saturdays.

5.4.1 Tissue Block or Slide Submission

At the time of biospecimen submission, a blank Clinical Assay Request form will be automatically printed at the same time as the standard BioMS shipping manifest. This form should be completed by appropriate clinical personnel and included with the biospecimen shipment to Cancer Genetics. It is needed, independent of the BioMS shipping form, for returning EGFR, ALK and PD-L1 testing results to the clinical site. Please also send a copy of the patient's pathology report to Cancer Genetics. This pathology report and Clinical Assay Request will NOT be forwarded to the BCR.

All tissue specimens for ALK, EGFR and PD-L1 analysis should be sent to Cancer Genetics:

Cancer Genetics
Attn: Clinical Trials
201 Route 17 North, 2nd Floor
Rutherford, NJ 07070
Tel: 201-528-9200

Once Cancer Genetics has analyzed tissue from the tissue block they will forward it to the NCI BCR for genomics. Sites submitting unstained slides to Cancer Genetics will also need to submit scrolls and an H&E slide to the NCI BCR for genomics.

5.4.2 Scrolls and H&E Slides Submission:

All scrolls and H&E slides (for those sites submitting slides to Cancer Genetics and scrolls to BCR) are to be submitted to:

NCI CCG Biospecimen Core Resource
Nationwide Children's Hospital
700 Children's Dr., WA1340
Columbus, OH 43205
Tel: 614-355-3589
ALCHBCR@nationwidechildrens.org

The BCR will bank tissue for future studies.

5.5 Tissue Submission at Progression

Patients undergoing a diagnostic biopsy at recurrence will have tissue submitted to the BCR, if available. Sites should submit FFPE material at room temperature (block preferred, scrolls with an H&E slide or unstained slides are acceptable) to the address below. The material should be labeled with the A151216 patient ID and the protocol ID (A151216). Please contact the BCR for questions on how to process FFPE scrolls or unstained slides.

If there were clinical genomics done on a recurrence specimen for any patient, regardless of whether or not there is recurrence tissue available for submission, sites should submit the molecular report on the recurrence specimen in RAVE.

Recurrence biopsies are to be submitted to:

NCI CCG Biospecimen Core Resource
Nationwide Children's Hospital
700 Children's Dr., WA1340
Columbus, OH 43205
Tel: 614-355-3589
ALCHBCR@nationwidechildrens.org

The BCR will bank the tissue for future studies.

5.6 Specimen Analysis

5.6.1 EGFR and ALK Genotyping

All patients (including those with local genotyping results) will have formalin-fixed tissue collected for central genotyping. The testing will be performed at Cancer Genetics, a commercial CLIA-certified laboratory. For non-squamous patients, ALK FISH will be performed using the Vysis break-apart probe and EGFR genotyping will be performed by sequencing of exons 18-21. EGFR and ALK genotyping results are expected to be provided to the treating clinician within 14 business days of submission so they can be used to determine eligibility for the randomized adjuvant studies, or to confirm the local results. Results will also be reported at intervals to the study team for upload into the Alliance database.

5.6.2 Next Generation Sequencing

In addition to the commercial genotyping at Cancer Genetics, tissue will be collected for research genomics by CCG. For those patients with a block available, this will be forwarded to the CCG after clinical testing at Cancer Genetics. For those patients without a block available, 10 micron scrolls should be cut from a block and submitted (the thicker sections reduce oxidative tissue damage seen with standard thickness slides; please see the scroll calculator on the Alliance and CTSU website for the number of scrolls to be cut). A peripheral blood specimen will also be collected and sent to the CCG BCR to be used as a source of non-malignant ('germline') DNA. Specimens will be coded. Over the course of the study, the CCG will perform advanced genomic analysis of the resected lung cancer specimens in a research, non-CLIA environment. Following completion of the genomic analysis, the results can be matched with the clinical follow-up results using a link between the samples coded and the patient identifiers for correlative analyses. The results of these genomic studies will not be provided back to the patient or their treating physician.

All remaining tissue will be stored at the CCG BCR for future studies.

5.6.3 PD-L1 IHC

To assess the role of PD-L1 protein expression as a biomarker, tumor tissue will be collected prospectively from all patients in this study. PD-L1 expression will be evaluated using the validated Dako 223 PD-L1 IHC assay. PD-L1 protein expression in NSCLC is determined by using Tumor Proportion Score (TPS), which is the percentage of viable tumor cells showing partial or complete membrane staining at any intensity. The specimen should be considered to have PD-L1 expression if TPS $\geq$ 1% and high PD-L1 expression if TPS $\geq$ 50%.

5.6.4 Genomic Analysis of Cell-free Plasma DNA

Two blood specimens will be collected for analysis of plasma cfDNA. Blood will be spun into plasma following receipt at the BCR. Plasma will be frozen until time of analysis, at which time DNA will be extracted from one specimen for genomic analysis; the second plasma specimen will be saved for future confirmatory studies. Over the course of the study, the CCG will coordinate advanced genomic analysis of the resected lung cancer specimens in a research, non-CLIA environment. Following completion of the genomic analysis, the results can be matched with the clinical follow-up results using a link between the samples coded and the patient identifiers for correlative analyses. The results of these genomic studies will not be provided back to the patient or their treating physician. A third blood specimen, collected in EDTA, will be used as a germline control for the planned tumor genomics.

5.7 Test Results from Cancer Genetics

Results will be sent to sites by FAX and to the e-mail address provided at the time of registration.

Squamous Patients: PD-L1 results will be returned to the sites within 7 business days of receipt.

Non-squamous Patients:

- o **For patients with no adjuvant therapy: EGFR, ALK and PD-L1 testing** will be performed and results will be returned to the sites within 14 business days.
- o **For patients who have received adjuvant therapy: EGFR and ALK testing** will be performed and results will be returned to the sites within 14 days of receipt.

If sites have questions about a delayed test result contact Cancer Genetics at clinicaltrials@cgix.com.

Once Cancer Genetics has completed testing on tissue blocks they will send the remaining block to the BCR for genomic analyses and storage.

5.8 Inadequate Submissions

5.8.1 Cancer Genetics

If the blocks or slides submitted to Cancer Genetics for molecular testing are inadequate, or fail to yield a result, Cancer Genetics will contact the site requesting an additional submission.

5.8.2 Biospecimen Core Resource

If the remaining tissue from Cancer Genetics or the scrolls submitted by sites are found to be inadequate for genomic analysis (DNA or RNA yield) the BCR will contact the site via e-mail and enter a specimen adequacy form into Medidata RAVE. If a site would like to submit additional specimens to BCR for genomic studies, please contact the BCR at: ALCHBCR@nationwidechildrens.org.

6.0 CLINICAL DATA REQUIREMENTS

6.1 Follow-up

- All patients NOT going onto an ALCHEMIST treatment trial will be followed on A151216 (every 6 months for 5 years). If a clinic visit did not occur within the 6 month window a site may call to follow the patient.

Patients that meet one or more of the following criteria will not be followed on A151216:

- Patients that are pre-registered only. This will be a very small percentage of patients who are registered pre-operatively and are found at surgery not to be eligible to participate on A151216.
- Patients enrolled on a treatment trial will be followed on the respective trial.

6.2 Data Collection and Submission

6.2.1 Data Submission Schedule

A Data Submission Schedule (DSS) is available on the Alliance study webpage, within the Case Report Forms section. The Data Submission Schedule is also available on the CTSU site within the study-specific Case Report Forms folder.

6.2.2 Medidata Rave

Medidata Rave is a clinical data management system being used for data collection for this trial/study. Access to the trial in Rave is controlled through the CTEP-IAM system and role assignments. To access Rave via iMedidata:

- Site staff will need to be registered with CTEP and have a valid and active CTEP-IAM account; and

- Assigned one of the following Rave roles on the relevant Lead Protocol Organization (LPO) or Participating Organization roster at the enrolling site: Rave CRA, Rave Read Only, Rave CRA (LabAdmin), Rave SLA, or Rave Investigator. Refer to https://ctep.cancer.gov/investigatorResources/default.htm for registration types and documentation required.
   - To hold Rave CRA or Rave CRA (Lab Admin) role, site staff must hold a minimum of an AP registration type;
   - To hold Rave Investigator role, the individual must be registered as an NPIVR or IVR; and
   - To hold Rave Read Only role, site staff must hold an Associates (A) registration type.

Upon initial site registration approval for the study in Regulatory Support System (RSS), all persons with Rave roles assigned on the appropriate roster will be sent a study invitation e-mail from iMedidata. To accept the invitation, site staff must log in to the Select Login (https://login.imedidata.com/selectlogin) using their CTEP-IAM username and password, and click on the accept link in the upper right-corner of the iMedidata page. Site staff will not be able to access the study in Rave until all required Medidata and study specific trainings are completed. Trainings will be in the form of electronic learnings (eLearnings), and can be accessed by clicking on the link in the upper right pane of the iMedidata screen. If an eLearning is required and has not yet been taken, the link to the eLearning will appear under the study name in iMedidata instead of the Rave EDC link; once the successful completion of the eLearning has been recorded, access to the study in Rave will be granted, and a Rave EDC link will display under the study name.

Site staff who have not previously activated their iMedidata/Rave account at the time of initial site registration approval for the study in RSS will also receive a separate invitation from iMedidata to activate their account. Account activation instructions are located on the CTSU website in the Rave section under the Rave resource materials (Medidata Account Activation and Study Invitation Acceptance). Additional information on iMedidata/Rave is available on the CTSU members' website in the Data Management > Rave section at www.ctsu.org/RAVE/ or by contacting the CTSU Help Desk at 1-888-823-5923 or by e-mail at ctsucontact@westat.com.

6.1.3   Data Quality Portal

The Data Quality Portal (DQP) provides a central location for site staff to manage unanswered queries and form delinquencies, monitor data quality and timeliness, generate reports, and review metrics.

The DQP is located on the CTSU members' website under Data Management. The Rave Home section displays a table providing summary counts of Total Delinquencies and Total Queries. DQP Queries, DQP Delinquent Forms and the DQP Reports modules are available to access details and reports of unanswered queries, delinquent forms, and timeliness reports. Review the DQP modules on a regular basis to manage specified queries and delinquent forms.

The DQP is accessible by site staff that are rostered to a site and have access to the CTSU website. Staff that have Rave study access can access the Rave study data using a direct link on the DQP.

To learn more about DQP use and access, click on the Help icon displayed on the Rave Home, DQP Queries, and DQP Delinquent Forms modules.

Note: Some Rave protocols may not have delinquent form details or reports specified on the DQP. A protocol must have the Calendar functionality implemented in Rave by the Lead Protocol Organization (LPO) for delinquent form details and reports to be available on the DQP. Site staff should contact the LPO Data Manager for their protocol regarding questions about Rave Calendaring functionality.

7.0   STATISTICAL CONSIDERATIONS

This is a central biomarker screening trial that is designed to screen resected lung cancers for targetable genomic alterations.

7.1   Sample Size

It is estimated that up to 8000 patients may need to be genotyped in order to fully accrue to the *EGFR* (estimated prevalence in advanced disease setting is 15%) and *ALK* (estimated prevalence in advanced disease setting is 5%) studies. We anticipate approximately 10% of patients screened in the adjuvant setting to have either the EGFR mutation or the ALK rearrangement. Fewer patients may need to be screened depending upon adoption of pre-screening strategies such as clinical selection or local genotyping.

It is estimated that up to 1000 patients with (-) EGFR and ALK non-squamous NSCLC and an additional 300 squamous cell NSCLC patients will be tested for PD-L1 to complete enrollment on EA5142. This assumes 20-25% of patients enrolled to EA5142 will have squamous cell carcinoma.

Thus, up to 8300 patients will be accrued to this screening trial to facilitate accrual to the three adjuvant trials: A081105, E4512 or EA5142.

7.2 Baseline Clinical Information

At time of registration, patients will assist the CRA in completing the on-study forms, to report characteristics that may be associated with the planned genomic analysis, including the following data:

- Age, gender, racial background
- Personal history of cancer and other pulmonary disorders
- Family cancer history (including family smoking history)
- History of occupational and environmental exposures including prior radiation
- Smoking history, including second hand smoke exposure

7.3 Clinical Follow-up Plan

Given the large number of subjects being followed, clinical follow-up will be kept to a minimum. All patients will be followed until otherwise notified by the Statistical Center. Patients will be contacted every 6 months to assess the following datapoints:

- Adjuvant therapy received (Y/N, which agents)
- Recurrent (Y/N)
- Date of recurrence
- Site of recurrence
- Pathologic confirmation of recurrence (Y/N, type of biopsy)
- Smoking Status
- Dead (Y/N)
- Date of death

In the instances where BCR has determined that there is not usable tissue for genomic analysis, the Data Center will contact the site to let them know patient follow-up may be discontinued. Sites should not discontinue patient follow-up before the 5-year point unless instructed to do so by the Data Center.

7.4 Endpoints

7.4.1 Primary Endpoint

There are two primary endpoints to this trial:

- Central clinical genotyping to facilitate accrual to the adjuvant Intergroup studies as measured by rate of accrual.

- Feasibility of research grade FFPE tissue collection for CCG analysis, as measured by adequate specimens collected per month. The goal is to achieve a collection rate over 100 adequate cases per month, to allow collection of at least adequate 4800 specimens over a four-year period. Importantly, this collection rate will depend upon specimen adequacy reports provided by the CCG.

7.4.2 Secondary Endpoints

There are two secondary endpoints for this trial:

- **2-year disease free survival (DFS) rate for lung cancers which are wild-type for EGFR and ALK.** Using genomics performed at CCG, DFS rate will be calculated for each genotype-defined population constituting greater than 1% of the study cohort. DFS is defined as the time from resection to the earliest of documented disease recurrence confirmed by biopsy, development of a new lung cancer confirmed by biopsy, or death from any cause. We estimate at least 80 patients in each of these rare genotype-defined subsets, which will allow estimation of the 2 year DFS rate within 11.2% points with 90% confidence. This will serve as a historical control for future single-arm phase II

trials of targeted adjuvant agents in these populations. Note: Patients enrolled in adjuvant targeted therapy trials, such as the adjuvant nivolumab trial (EA5142), will be excluded from this estimate of DFS.

- **Agreement of local genotyping methods (direct sequencing of EGFR, ALK FISH) with central CLIA genotyping.** Each locally deemed EGFR-mutant or wild-type patient will also be classified by central assessment. Similarly, each patient deemed locally as ALK-rearranged or not by FISH will be classified by the central assessment. For each locally used assay, agreement will be defined as the proportion of patients deemed mutant (or wild-type) by local and central assessment divided by the number of evaluable patients, where an evaluable patient is one who has a local assessment result and has submitted tissue for central assessment. An agreement rate of 90% or higher between the local assay and the central assessment will be deemed acceptable. The 95% confidence intervals for 90% success rates such that the lower limit is at least 80% or higher are given in the following table for different sample sizes:

Sample Size	Number of successes	95% Confidence Interval (lower, upper)
100	90	82.4, 95.1
150	135	84.0, 94.3
200	180	85.0, 93.8
250	225	85.6, 93.4

7.4.3 Exploratory/Other Endpoints

- **Spectrum of new mutations identified at recurrence.** Genomic analysis will be performed on tissue collected at time of recurrence and compared to baseline genomics. New mutations in key oncogenes and tumor suppressor genes (PIK3CA, PTEN, etc) will be quantified. It is hypothesized that a greater number of new alterations will be identified in patients whom received adjuvant chemotherapy as opposed to those not receiving adjuvant chemotherapy.

- The reasons behind why potentially eligible ALK-rearranged/EGFR mutant patients decline to enroll onto the adjuvant trials will be summarized. Specifically, the proportions of patients who decline to enroll because of concern with randomization, or not needing further therapy versus those who become otherwise ineligible due to recurrent disease or missing the enrollment window will be catalogued. Such summaries will be periodically reviewed by the study team to understand if any changes or clarifications are needed to the protocol or if additional educational material needed for the sites to help facilitate accrual to the adjuvant studies.

- Simple exploratory analyses will be used to better understand the variability in the levels of baseline cfDNA based on the timing of collection of these samples. The levels of cfDNA (stratified by the timing of collection) will be correlated with clinical outcomes of overall survival and disease-free survival using Kaplan-Meier approach as well as exploratory Cox proportional Hazards models adjusted for baseline smoking, patient, and tumor characteristics as well as treatment information.

7.5 Sample Size

The sample size will depend partially upon the prevalence of EGFR mutations and ALK rearrangements, and partially upon the degree of selection used when investigators are accruing patients. With no clinical selection, up to 8300 patients will need to be screened to fully accrue the randomized adjuvant studies. This is because ALK rearrangements are present in 4-5% of lung adenocarcinoma, such that 8000 patients must be screened to identify the 366 subjects for the crizotinib study. However, if investigators decide to use clinical selection methods to

determine which patients to screen, and primarily accrue never-smokers, then half as many patients must be screened (EGFR mutations and ALK rearrangements are twice as prevalent in never-smokers as in the general adenocarcinoma population). Alternatively, some centers may genotype resected cancers locally and then accrue patients with a known EGFR mutation or ALK rearrangement – this would further decrease the total number of patients needed to be centrally genotyped to achieve the study aims.

An additional cohort of 300 patients with resected squamous cell carcinoma will be enrolled for the purposes of PD-L1 screening and for assessment to be enrolled on EA5142.

7.6   Analysis Plan

Accrual rate to the adjuvant studies will be monitored every 3 months, and discussed between the study teams coordinating the ALCHEMIST study and the adjuvant studies. If accrual is inadequate, then the ALCHEMIST study will initiate strategies to improve accrual, including opening the screening study at new centers and developing strategies for genotyping at participating centers to improve catchment. Specimen collection rate will also be monitored every 3 months and discussed between the ALCHEMIST study team and the CCG. If collection of adequate specimens is insufficient, then the ALCHEMIST study will initiate strategies to improve specimen adequacy.

7.7   Inclusion of Women and Minorities

This study will be available to all eligible patients, regardless of race, gender, or ethnic origin.

Accrual Targets			
Ethnic Category	**Sex/Gender**		
	Females	**Males**	**Total**
Hispanic or Latino	129	166	295
Not Hispanic or Latino	4482	3523	8005
Ethnic Category: Total of all subjects*	4611	3689	8300
Racial Category			
American Indian or Alaskan Native	19	19	38
Asian	0	0	0
Black or African American	239	185	424
Native Hawaiian or other Pacific Islander	0	0	0
White	4352	3486	7838
Racial Category: Total of all subjects	4610	3690	8300

Ethnic Categories: **Hispanic or Latino –** a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. The term "Spanish origin" can also be used in addition to "Hispanic or Latino."
Not Hispanic or Latino

Racial Categories:	**American Indian or Alaskan Native –** a person having origins in any of the original peoples of North, Central, or South America, and who maintains tribal affiliations or community attachment. **Asian –** a person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. (Note: Individuals from the Philippine Islands have been recorded as Pacific Islanders in previous data collection strategies.) **Black or African American –** a person having origins in any of the black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American." **Native Hawaiian or other Pacific Islander –** a person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. **White –** a person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

8.0   ETHICAL AND PRIVACY CONSIDERATIONS

8.1   Restricted Access to Genomic Data

The research genomic studies will generate genetic data unique to an individual ("genetic fingerprints", or genotypes). These data are not directly tied to an identified individual, and the clinical information associated with these data will be de-identified. Nevertheless, a risk exists that the genetic data could lead to the re-identification of a participant or relative. Consequently, NIH policy is that individual genetic data from the characterization studies are kept in a restricted-access tier of the database.

To be authorized to access the restricted tier of data, Investigators must submit an application to a Data Access Committee (DAC) of the National Institutes of Health designated to review applications for the Alchemist initiative. Upon approval by the DAC that the access request is for bona fide research purposes, the Investigator, scientists under their control, and their institution must subscribe to a Data Use Certification (DUC) that controls their ability to access the data, redistribute the data, prohibits the re-identification of participants, and includes requirements for data security. Controlled-access data are for General Research Use, i.e. usable for any genetic studies; there are no data use restrictions with respect to field of study and users may apply data to any legitimate research including non-cancer research-related discovery.

9.0 REFERENCES

1. Rosell R, Moran T, Queralt C, Porta R, Cardenal F, Camps C, et al. Screening for epidermal growth factor receptor mutations in lung cancer. N Engl J Med. 2009;361:958-67.

2. Kwak EL, Bang Y-J, Camidge DR, Shaw AT, Solomon B, Maki RG, et al. Anaplastic Lymphoma Kinase Inhibition in Non–Small-Cell Lung Cancer. N Engl J Med. 2010;363:1693-703.

3. Richardson F, Richardson K, Sennello G, Young D, Orlov S, Papai-Szekely Z, et al. Biomarker analysis from completely resected NSCLC patients enrolled in an adjuvant erlotinib clinical trial (RADIANT). ASCO Meeting Abstracts. 2009;27:7520.

4. Goss GD, O'Callaghan C, Lorimer I, Tsao M-S, Masters GA, Jett J, et al. Gefitinib Versus Placebo in Completely Resected Non–Small-Cell Lung Cancer: Results of the NCIC CTG BR19 Study. J Clin Oncol. 2013.

5. Johnson BE, Kris MG, Berry LD, Kwiatkowski DJ, Iafrate AJ, Varella-Garcia M, et al. A multicenter effort to identify driver mutations and employ targeted therapy in patients with lung adenocarcinomas: The Lung Cancer Mutation Consortium (LCMC). ASCO Meeting Abstracts. 2013;31:8019.

6. Brahmer J, Reckamp KL, Baas P, Crino L, Eberhardt WE, Poddubskaya E, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. N Engl J Med. 2015; 373: 123-35.

7. Paz-Ares L, Horn L, Borghaei H, Spigel D, Steins M, Ready N, et al. Phase III, randomized trail of nivolumab versus docetaxel in advanced non-squamous cell non-small cell lung cancer. J Clin Oncol. 2015; 33: LBA 109.

8. Garon EB, Rizvi NA, Hui R, Leighl N, Balmanoukian AS, Eder JP, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. N Engl J Med. 2015; 372: 2018-28.

APPENDIX I

<u>For laboratory use: ALCHEMIST laboratory flow diagram</u>
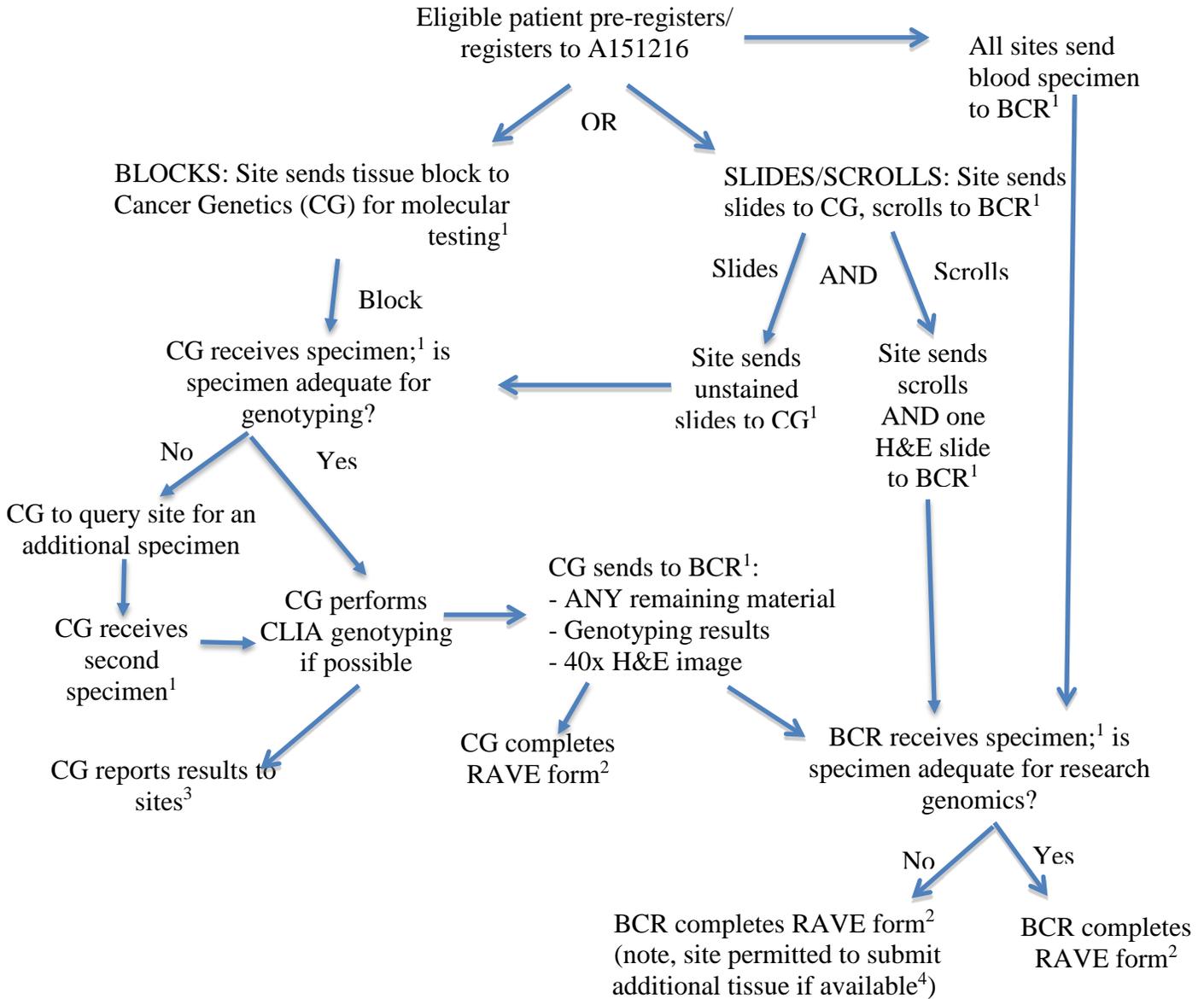
[1]All specimens sent or received must be logged into BioMs.

[2]Sites will have view access to these results & adequacy CRFs. Alliance Data Center will review specimen adequacy results at least quarterly to identify problems and implement improvements.

[3]EGFR/ALK and PD-L1 results will be returned by FAX to the CRA and ordering clinician using contact information from requisition. The EGFR/ALK results are sent within 14 business days of receipt. In squamous patients PD-L1 testing will be returned within 7 business days of receipt; PD-L1 results will be returned within 21 business days in the EGFR/ALK (-) patients.

[4]Sites with inadequate specimens will be contacted by the BCR to discuss resubmission.

APPENDIX II

Laboratory Manual for ALK/EGFR Testing Tissue Submission ALCHEMIST

CANCER GENETICS CONTACT INFORMATION:

Shipping Address:
Cancer Genetics, Inc
Attn: Clinical Trials
201 Route 17 North, 2nd Floor
Rutherford, NJ 07070
clinicaltrials@cgix.com
Tel: 201-528-9243

BLOCK AND SLIDE PREPARATION

PREPARATION OF PARAFFIN EMBEDDED TISSUE BLOCKS

Tissue blocks are preferred. Blocks should be submitted as soon as possible after consent is obtained.

- Standard dimensions of block – approximately 4cm x3cm
- If multiple blocks available, **submit block with most tissue**; do **not** send multiple blocks per subject.
- The subject identifier and block number must be written on the block in pencil and be clearly legible.
- Wrap blocks in a foam pouch and place into slide container.
- See packaging instructions.

PREPARATION OF TISSUE SLIDES

If the tissue block is unavailable, prepare tissue slides.
- **The number of slides to be submitted is** five (5) 10-micron sections plus eight (8) 5-micron sections.
- **Positively charged frosted ended slides must be used.**
- Section a single section containing tissue onto each slide.
- Sections **must** be from the same tissue block.
- Subject identifier, and block number must be written legibly in pencil on the frosted end of the slide.
- Slides must **not** be baked or melted.
- Cover slips must **not** be used.
- Sections must **not** be stained.
- Place slides in slide box.
- See packaging instructions.

IMPORTANT HIGHLIGHTS ABOUT TISSUE SAMPLE PREPARATION
- Tissue blocks are preferred.
- Multiple blocks per visit should not be submitted.
- Tissue sections **must** come from the same block.
- The entire clinical assay request form must be completed and include the biopsy collection date and the biopsy collection site. Identifiers including subject identifier, and block ID number must be entered on the requisition form and must match the identifiers on the tissue blocks or slides.
- Only the required number of slides or a block of tissue should be submitted.

AVOID THE FOLLOWING
- Submission of the incorrect number of slides
- Submission of stained slides
- Incorrect sample packaging
- Incorrect and/or incomplete clinical assay request form (e.g. Subject identifier and block number missing from the block or slides)

OVERVIEW OF TUMOR BLOCK SHIPPING

1. Fill out the Clinical Assay Request Form and place in the foam mailer.	2. Ensure the tissue block is labeled legibly with the Subject Identifier and block number, written in pencil.	3. Place the tissue block in a slide box.
4. Place the slide box in the foam mailer.	5. Ship ambient to Cancer Genetics Shipping Address: Cancer Genetics, Inc Attn: Clinical Trials 201 Route 17 North, 2nd Floor Rutherford, NJ 07070 Tel: 201-528-9200 clinicaltrails@cgix.com	

OVERVIEW OF SLIDE SHIPPING

1. Five (5) newly cut serial tissue sections cut at 10-micron plus eight (8) cut at 5-micron are mounted on positively charged frosted ended slides.	2. Ensure that subject identifier and block number are written legibly in pencil on the frosted end of each slide.	3. Place the slides in a slide box.
4. Place the slide box and the Clinical Assay Request Form in a foam mailer.	5. Ship ambient to Cancer Genetics Shipping Address: Cancer Genetics, Inc Attn: Clinical Trials 201 Route 17 North, 2nd Floor Rutherford, NJ 07070 Tel: 201-528-9200 clinicaltrails@cgix.com	

APPENDIX III: GENOMIC ANALYSIS PLAN

Introduction

A151216 is an ongoing prospective trial led by the Alliance for Clinical Trials in Oncology and sponsored by CTEP. Patients are enrolled following resection of high risk early stage non-small cell lung cancer (NSCLC). The study has two parallel aims. First, clinical genotyping is performed at a core laboratory to facilitate enrollment to randomized trials of adjuvant targeted therapies; enrollment is ongoing with on target enrollment to at least some of these studies. Second, collection of tissue and clinical data is performed to permit clinically-informed genomic analysis of resected NSCLC; this document is intended to structure the genomic analysis described in the second aim of A151216.

As of April 1 2019, 4409 patients (Cohort A) have been enrolled to A151216 from 678 sites in the US, noting that enrollment continues at a rate exceeding 100 patients monthly so all enrollment numbers are quickly outdated. Of these 4409 patients, 1125 have enrolled onto one of the ongoing adjuvant treatment trials, for these patients clinical follow-up has been performed under the separate treatment trial and is DSMB protected, thus leaving 3284 patients (Cohort B). Adequate FFPE tissue and blood has been received to permit genomic analysis from 2070 patients (Cohort C). Study follow-up continues to capture recurrence and death, for a maximum of 5 years following enrollment. See **Figure 1** for details.

Figure 1: Description of the cohorts available for the different aims
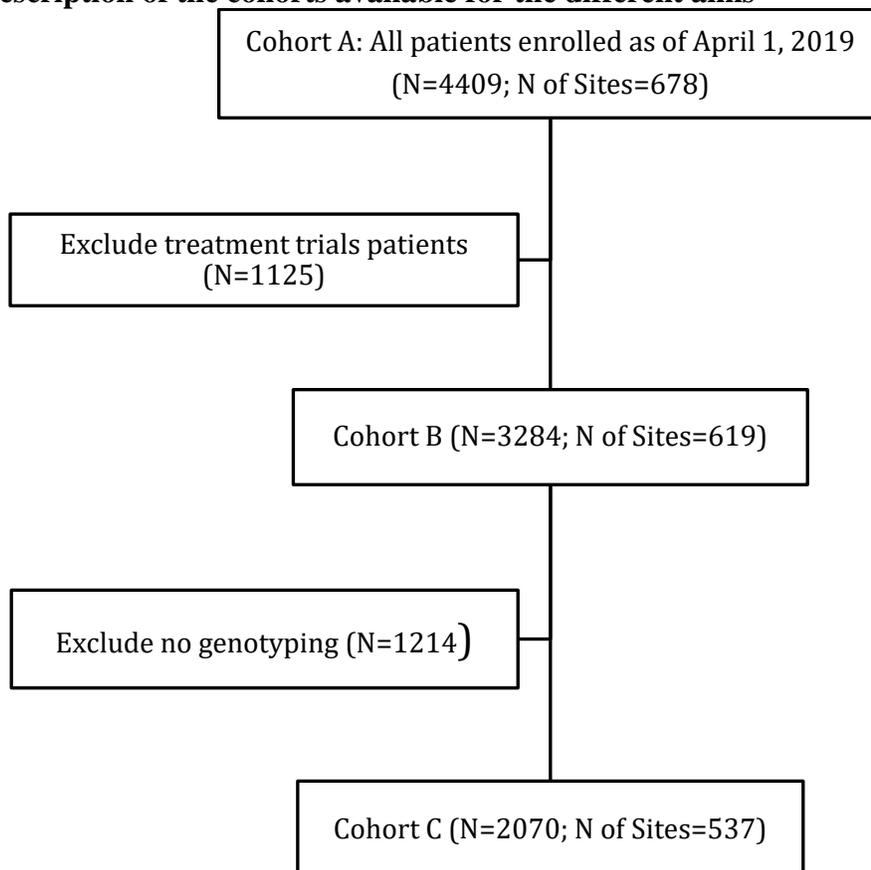
Rationale

The potential learnings from these large, clinically annotated cohorts are extensive. Here we propose to focus on three leading areas of investigation, understanding that additional questions will arise during the analysis:

Disparities in the care of resected NSCLC: On a population basis, many patients with lung cancer do not receive a full course of standard therapy. Patients with early-stage disease commonly do not undergo surgical resection, and many patients with advanced disease do not receive systemic treatment or guideline-concordant molecular profiling. Disparities in treatment also exist according to socioeconomic factors, race, and geography, which may contribute to disparities in clinical outcomes. To improve care quality across the population of patients with lung cancer, it is critical to understand the non-clinical factors that contribute to differences in treatment. The ALCHEMIST study presents a unique opportunity to study this variation in lung cancer treatment, since it is conducted across the National Clinical Trials Network (NCTN), which includes both academic and community-based sites.

Genomics of resected NSCLC: The large size of the A151216 cohort and extensive clinical and genomic annotation makes this a rich cohort for genomic discovery efforts. These include opportunities to study the heterogeneity of existing genomic subtypes of NSCLC (e.g. KRAS-mutant, EGFR-mutant) as well as to discover new, rare genomic subtypes of NSCLC (e.g. novel oncogenic fusions). While there exists extensive genomic data describing advanced NSCLC, this study offers a unique opportunity to study how the genomic make-up of early stage cancer differs from that of advanced cancer.

Prognostic biomarkers for resected NSCLC: Across all ALCHEMIST treatment arms, enrollment has been challenged in part because patients are disinclined to receive additional treatment after enduring the toxicities of standard adjuvant chemotherapy. One candidate solution, for appropriate patients, could be to intensify adjuvant therapy immediately after surgery – adding targeted therapy to chemo. The success of such an aggressive approach would be greatly strengthened by biomarkers to better identify patients at highest risk of recurrence, patients who would be motivated to intensify adjuvant therapy. We believe that the ubiquitous nature of NSCLC genomic analysis makes it a highly attractive and scalable clinical diagnostic which could be used to identify patients at highest risk of recurrence after resection.

Stakeholders

A151216 relies upon the contributions of a large team of clinicians and scientists. The planned analysis is expected to involve participation by representatives from the NCI Cancer Therapy Evaluation Program (CTEP), including the Alliance for Clinical Trials in Oncology (e.g. Respiratory Committee and the Statistical and Data Center (SDC)) and the National Clinical Trials Network (e.g. ECOG-ACRIN and SWOG), as well as representatives from the NCI Center for Cancer Genomics (CCG), including the Biospecimen Core Resource (BCR), the Genome Characterization Centers (GCC), the Genomic Data Analysis Network (GDAN). To allow these analyses proposed in this amendment, an Analysis Working Group including these stakeholders will be assembled made up of representatives from these various stakeholders.

Genomic analysis

The genomic analysis of specimens collected on A151216 is implemented by CCG. All patients enrolled to A151216 are expected per protocol to submit FFPE tissue and whole blood to the BCR. FFPE tissue is either submitted as a block or tissue scrolls. This tissue is processed, and specimens deemed adequate are then sequenced as follows:

Whole exome sequencing at the Broad Institute
Library construction is performed as described previously (Fisher et al, Genome Biology, 2011), with the following modifications: initial genomic DNA input into shearing is reduced from 3µg to 10-100ng in 50µL of solution. For adapter ligation, Illumina paired end adapters are replaced with palindromic forked adapters, with unique 8 base molecular barcode sequences included in the adapter sequence to facilitate downstream pooling. With the exception of the palindromic forked adapters, the reagents used for end repair, A-base addition, adapter ligation, and library enrichment PCR were purchased from KAPA Biosciences in 96-reaction kits. In addition, during the post-enrichment SPRI cleanup, elution volume was reduced to 30µL to maximize library concentration, and a vortexing step was added to maximize the amount of template eluted. After library construction, hybridization and capture are performed using the relevant components of Illumina's Rapid Capture Exome Kit and following the manufacturer's suggested protocol, with the following exceptions: first, all libraries within a library construction plate are pooled prior to hybridization. Second, the Midi plate from Illumina's Rapid Capture Exome Kit is replaced with a skirted PCR plate to facilitate automation. All hybridization and capture steps are automated on the Agilent Bravo liquid handling system.

After post-capture enrichment, library pools are quantified using qPCR (automated assay on the Agilent Bravo), using a kit purchased from KAPA Biosystems with probes specific to the ends of the adapters. Based on qPCR quantification, libraries are normalized to 2nM, then denatured using 0.1 N NaOH on the Hamilton Starlet. After denaturation, libraries are diluted to 20pM and then loaded across the appropriate number of Illumina flowcell lanes to achieve the target coverage.

Whole genome sequencing at the Broad Institute
An aliquot of genomic DNA is taken from a stock sample at a target of 350ng in 50µL of solution to serve as the input into shearing. Samples undergo fragmentation by means of acoustic shearing using Covaris focused-ultrasonicator, targeting 385bp fragments. Following fragmentation, additional size selection is performed using a SPRI cleanup. Library preparation is performed using a commercially available kit provided by KAPA Biosystems (KAPA Hyper Prep without amplification module, product KK8505), and with palindromic forked adapters with unique 8 base index sequences embedded within the adapter (purchased from IDT). Following sample preparation, libraries are quantified using quantitative PCR (kit purchased from KAPA biosystems) with probes specific to the ends of the adapters. This assay was automated using Agilent's Bravo liquid handling platform. Based on qPCR quantification, libraries are normalized to 1.7nM. Samples are then pooled into 24-plexes and the pools are once again qPCRed and then loaded across the appropriate number of Illumina flowcell lanes to achieve the target coverage.

Total RNA sequencing at the University of North Carolina
Libraries for fresh frozen (FF) samples and formalin-fixed paraffin-embedded (FFPE) samples are prepared in batches of 24 to 96 (increasing in units of 8) using the TruSeq Stranded Total RNAseq library kit with Ribo-Zero Gold depletion on a Bravo Liquid Handling System.

After initial receipt, samples are evaluated for RNA quality. Samples are transferred from Matrix tubes to skirted 96-well plates. A small aliquot is removed and stored for downstream genotyping, if needed. FF samples are assessed for concentration of RNA, RIN score, and, as needed, fragment size using a TapeStation 2200 (Agilent, Santa Clara, CA) or a plate-reader. FFPE samples are assessed for concentration of RNA and fragment size using a TapeStation 2200.

Qualifying samples are processed through the TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold High Throughput (96 samples, 96 indexes catalog# RS-122-2303, Illumina) protocol. This procedure is largely performed on Agilent Bravo automated Liquid handling Platform (Model 16050-102) using custom automation scripts. In this procedure, ribosomal RNA (rRNA) is removed using biotinylated, target-specific oligos in conjunction with Ribo-Zero rRNA removal beads. Following purification, the FF RNA is incubated with divalent cations at elevated temperature to fragment the RNA into small pieces. This step is typically not needed for FFPE samples as these are usually fragmented as a result of the formalin-fixation. The first strand of cDNA is synthesized from the RNA fragments using reverse transcriptase and random primers. Spurious DNA-dependent synthesis is chemically inhibited, which improves strand specificity. After cleanup, second strand synthesis is performed using DNA Polymerase I and RNase H. Replacement of dTTP with dUTP in the reaction results in strand specificity, as the incorporation of dUTP in second-strand synthesis quenches the second strand during subsequent amplification. The 3' ends are then adenylated to reduce the probability of multiple fragments being concatenated. Adaptors, including indices for multiplexing samples, are then ligated to the ends of the double-stranded cDNA molecules. Depending on project, either single-end indices or dual indices are used. Polymerase reaction is then used to amplify fragments. Samples are cleaned and selected for desired size.

Library concentration and fragment size are measured using a TapeStation. For FFPE samples, aliquots of passing samples are pooled into a high multiplex pool and sequenced with a Nano flowcell on an Illumina MiSeq sequencer (Illumina, San Diego, CA). Resulting sequences are demultiplexed using a custom quality-assessment bioinformatics pipeline. Data from each sample is evaluated to determine if the library is of sufficient quality to continue to HiSeq sequencing.
Libraries that qualify are pooled for sequencing on either an Illumina HiSeq 2500 or 4000 sequencer (Illumina, San Diego, CA); pools are diluted to around 250 pM, and loaded onto a flowcell following standard protocol. Samples are paired-end sequenced for 50–75 bp in pools of two to four samples, depending on the project. Sequence data is then post-processed to demultiplex indexed samples within the pool. Pools are bioinformatically evaluated for balance among samples and for total production. Individual samples are assessed for enrichment of coding RNAs and degree of rRNA contamination, as well as for a number of ancillary metrics. Samples that pass are then available for distribution.

miRNA sequencing at the University of North Carolina
Small RNA-seq libraries are prepared in batches of 8 to 96 samples using the Perkin Elmer NEXTflex Small RNAseq kit on a Perkin Elmer SciClone Liquid Handling System.

After initial receipt, samples are evaluated for RNA quality. Samples are transferred from Matrix tubes to skirted 96-well plates. Fresh frozen samples (FF) samples are assessed for concentration of RNA, RIN score, and, as needed, fragment size using a TapeStation 2200 (Agilent, Santa Clara, CA) or a plate-reader. Formalin-fixed paraffin-embedded (FFPE) samples are assessed for concentration of RNA and fragment size using a TapeStation 2200.

Libraries are constructed using the BioO NEXTflex Small RNA Sequencing Kit v3 for SciClone Automation (96 reactions) (PerkinElmer, catalog# NOVA-5132-08) on PerkinElmer SciClone G3 NGS Workstation Liquid Handling Robot. This protocol ligates an adenylated adaptor to the 3' end of the RNA molecules and then a non-adenylated adaptor to the 5' end of the RNA molecule. A primer complimenting the 3' adaptor primes a reverse transcription reaction; a primer complimenting the 5' adaptor then primes extension of the other strand. Products from this reaction are cleaned and size-selected using magnetic beads; samples are then PCR-amplified with universal primers and again cleaned by a gel-free method.

Libraries are assessed for quality using TapeStation or Qubit. The samples are then pooled according to desired depth of sequencing. Pools are size-selected for 130–170 base pairs using PAGE on a Sage Science Pippin Prep (Sage Science, Beverly, MA) via a 3% cassette, and their molarity is measured. Pools that qualify are sequenced on an Illumina HiSeq 4000 sequencer (Illumina, San Diego, CA); samples are diluted to around 250 pM and loaded onto a flowcell following standard protocol. Samples are then single-end sequenced for 50 bp. Sequence data is then post-processed to demultiplex indexed samples within the pool. Pools are bioinformatically evaluated for balance among samples and individual samples are assessed for enrichment of small RNAs; those that pass are then available for distribution.

Germline sequencing
Paired WGS and WES will be performed as part of the above tumor genomic analyses. This data will be available for study of previously described germline variants believed to impact germline risk (e.g. mutations in TP53, BRCA1/2, EGFR, HER2). Genome-wide germline discovery will not be a focus of the current analysis plan.

Tumor mutational burden
Tumor mutational burden (TMB) is the measure of the total number of nonsynonymous somatic mutations (missense mutation, nonsense mutation, and splice-site mutation) and short indels (in-frame deletion, in-frame insertion, frame-shift deletion, frame-shift insertion) per coding area of a tumor genome (i.e., size of the exome captured). We will discard germline variants and synonymous variants in the calculation of TMB as these variants are less likely to be directly involved in tumorigenesis and generating neoantigens. Specifically, germline variants can be filtered out using the matched non-tumor samples and the public databases such as dbSNP and 1000 Genomes. We will also filter out mutations with mutated allele frequency (MAF) less than 5% and a sequence depth less than 20X in tumor samples or less than 10X in normal samples.

Whole genome doubling
Ploidy abnormality, including whole genome doubling (WGD), is a hallmark of human cancers. Over 30% of NSCLC tumors have WGD [1]. We will use the ABSOLUTE algorithm to jointly estimate tumor purity and absolute copy number [2]. A WGD will be called if over 50% of the autosomal tumor genome with an MCN (major copy number, or the number of copies of the more frequent allele) of two or greater.

To get a tab-delimited segmentation file as input to ABSOLUTE, we will align the raw reads to the human reference genome (hg38/GRCh38) using BWA MEM [3] and reads coverage signals will be extracted from the BAM files to compute GC corrected probe level log2 ratios (tumor versus germline) using a modified version of PatternCNV [4]. Only the 3'UTR probes (which are only on the V4+UTR base kit) will be used to calculate copy numbers to avoid artifactual copy

number arising from stoichiometric variation in the dosage of the custom probes. We will then average the GC-corrected probe level log2 ratio to get the gene level log2 ratio, which will be segmented using the Bioconductor DNACopy version of CBS (https://bioconductor.org/packages/release/bioc/html/DNAcopy.html). Segmented log2 ratios will be rescaled to tumor purity so that all log2 levels are equivalent to a 100% pure tumor.

Additional exploratory genomic analyses

According to data from the TCGA Pan-Lung Cancer cohort, the vast majority (96%) of mutated genes have mutation frequencies < 5% [5]. Low mutation frequency would significantly reduce the statistical power of association analysis between genomic variants and clinical outcomes at the single gene level. To search for additional meta-genomic prognostic factors, we will perform association analysis at gene set (i.e. a group of biologically related genes) level. In other words, we will test for the association between clinical outcomes and an aggregated mutation residing in the selected gene set. We will use gene sets predefined in the mSigDB database (http://software.broadinstitute.org/gsea/index.jsp), including: 1) 50 "hallmark gene sets" that represent specific well-defined biological states and processes. 2) 5501 "curated gene sets" that were curated from various pathway databases (BIOCARTA, KEGG, PID, REACTOME) and the biomedical literature. 3) Gene Ontology gene sets. 4) Oncogenic gene sets that represent signatures of cellular pathways that are often dis-regulated in cancer.

Epidemiologic data

For each patient enrolled to A151216, the site study team documents clinical and pathologic characteristics as well as completing an epidemiologic survey. These key clinical features could impact the biology of the patient's lung cancer. These data fall into the following broad categories:
- Lung cancer stage and histology
- Past medical history
- Smoking history
- Occupational and environmental exposures
- Family history
- Other social history

These baseline clinical data are transferred on a monthly basis to BCR and are stored with the tumor sequencing data in preparation for the planned analyses.

Clinical follow-up-data

Clinical follow-up is performed for each patient enrolled onto A151216. Note that for those patients enrolled onto the therapeutic trials (e.g. erlotinib, crizotinib, immunotherapy), clinical follow-up is performed on those studies and is controlled by a Data Safety Monitoring Board (DSMB), therefore is unavailable for the investigations described here. The remaining patients are followed to capture recurrence, overall survival, and treatments delivered. These clinical follow-up data are housed at the Alliance SDC.

Aim 1: to understand patterns and determinants of standard-of-care curative therapy for resected high-risk NSCLC within the NCTN.

In this aim, we will measure the proportion of patients in the ALCHEMIST trial who received (a) any adjuvant chemotherapy, and, among those receiving any chemotherapy, (b) cisplatin-based treatment. We will then evaluate the proportion of variation in the utilization of (a) any adjuvant

chemotherapy, and (b) cisplatin-based treatment, that is explained by clinical vs non-clinical factors. A similar approach can also be employed to study other metrics of care quality for lung cancer, including anatomic surgical resection, adequacy of intraoperative lymph node dissection, and preoperative mediastinal staging.

Independent variables
- Clinical factors will include age, stage, histology, smoking history, and comorbidities.
- Non-clinical factors will include race, educational attainment, geography, and ALCHEMIST enrollment site category (NCTN-LAPS, NCORP, or other sites).

Analysis plan

The statistical analysis of these data will be performed by the Alliance SDC using the clinical and epidemiologic data collected on A151216 (Cohort B patients). Two multivariable logistic regression models will be constructed, following the principles outlined in the section below on the Management of Exploratory/Validation Cohorts. In the first, the outcome will be receipt of any adjuvant chemotherapy. In the second, the outcome will be receipt of any cisplatin-based chemotherapy, and Cohort B will be further subset to those patients who received any adjuvant chemotherapy. Independent variables will include those described above. In each case, the proportion of variance in the outcome explained by the clinical vs non-clinical factors will be calculated.

Aim 2: Describe the molecular epidemiology of resected NSCLC.

Prior reports from The Cancer Genome Atlas (TCGA) have offered striking insights into the broad range of genomic variants present within resected NSCLC. ALCHEMIST offers an opportunity to build upon these prior discoveries in a diverse clinical cohort representative of the US lung cancer population, with robust clinical annotation. This aim will include analysis of patient characteristics at time of enrollment as well as smoking and environmental exposure history, to permit the clinical features associated with a full spectrum of genomic characteristics.

Analysis plan

This analysis will be performed by the Alliance SDC, in conjunction with the A151216 Analysis Working Group , combining the genomic data contributed from the GDAN and the demographic data contributed from the Alliance SDC for Cohort C patients. The analysis will describe the prevalence of genomic changes detected on sequencing with a focus on rare genotypes as well as heterogeneity within common genotypes. Summative genomic features such as tumor mutational burden (TMB), and whole genome doubling (WGD) will also be studied, including a description of their relationship to common NSCLC genotypes (e.g. EGFR and KRAS mutations). Note that while intratumoral heterogeneity (ITH) is of interest, it cannot be calculated from the available data in the absence of multi-regional sampling. Prevalence of previously described germline risk alleles will also be reported. If the non-treatment trial cohort differs systematically from those on treatment trials, weighting will be considered so that inferences can be applied to the entire screening population. Thus, the inferences will apply to the entire ALCHEMIST screening population.

Sub-aim 2A: Detection of rare molecular subtypes of NSCLC

Estimating approximately 2000 cases available for genomic discovery efforts, we have exquisite statistical power to identify, and to characterize the prevalence of, even rare genomic events. Prevalence rates together with exact confidence intervals will be estimated.

Table 1: Confidence interval for estimating the prevalence of rare genotypes within the ALCHEMSIT cohort

Sample size	Number of Cases	Prevalence	Exact 95% CI (lower, upper)
2000	2	0.1%	0.01%, 0.36%
2000	4	0.2%	0.05%, 0.51%
2000	6	0.3%	0.11%, 0.65%
2000	8	0.4%	0.17%, 0.79%
2000	10	0.5%	0.24%, 0.92%
2000	12	0.6%	0.31%, 1.05%
2000	14	0.7%	0.38%, 1.17%
2000	16	0.8%	0.46%, 1.30%
2000	18	0.9%	0.53%, 1.42%
2000	20	1%	0.61%, 1.54%
2000	100	5%	4.09%, 6.05%
2000	200	10%	8.72%, 11.40%
2000	300	15%	13.46%, 16.64%
2000	400	20%	18.27%, 21.82%

Sub-aim 2B: Molecular epidemiology of molecular subtypes of NSCLC
Epidemiologic data collected as part of A151216 will be used to describe the epidemiologic characteristics associated with each genomic feature identified through sequencing. The analyses will follow the principles outlined in the section below on management of exploratory and validation cohorts. One key hypothesis is that presence or absence of mutation status is related to smoking history and other exposures, considering other characteristics (medical history, social history) as key covariate. A second key hypothesis is that family cancer history is associated with presence or absence of somatic and germline mutations, considering smoking history other exposures as key covariates. Finally, pathologic features (stage, grade) will be associated with mutation status, considering smoking history and other exposures as key covariates. Final estimates will be based on the development and validation cohorts combined for maximum possible precision. Logistic regression models will form the basis of this analysis.

Sub-aim 2C: Genomic profile of molecular subtypes of NSCLC
Broad sequencing efforts permit rich genomic description of molecular subtypes of NSCLC. This analysis will focus on established molecular subtypes of NSCLC, such as those positive for mutations in oncogenes like KRAS, EGFR, HER2, BRAF, ALK, RET, ROS, MET. The co-occurrence of other genomic events will be studied across these NSCLC subtypes, for example studying presence of mutations in TP53 and other tumor suppressor genes, or presents of relative gain or loss of chromosomal regions. Summative genomic features (TMB and WGD, as described above) will also be characterized across these NSCLC subtypes. Prevalence estimates with confidence intervals will be generated. Association with NSCLC subtypes or co-occurrence of genomic events will be assessed via contingency table methods.

Aim 3: Identify genomic features associated with early NSCLC recurrence.

Next-generation sequencing of NSCLC is increasingly standard for advanced disease but remains investigational for early stage NSCLC. One potential application of such an assay would be for identification of patients at high risk of recurrence, who might have potential to benefit from an

intensified adjuvant treatment approach. ALCHEMIST will permit study of genomic features associated with risk of early recurrence in a large cohort (Cohort C) that will also allow for multivariable analysis of known prognostic variables (e.g. stage, receipt of adjuvant chemotherapy).

Recent data from the TRACERx program has revealed important genomic features indicating a poor prognosis for resected NSCLC [6]. Studying 100 cases of resected NSCLC, the investigators performed multi-regional sampling and WES to assess ITH. Copy number heterogeneity, as measured by the percent of copy number alterations that were subclonal, was found to be highly associated with risk of recurrence (HR 4.9, $p=4.4 \times 10^{-4}$), and remained significant upon multivariate analysis (p=0.01). Importantly, one of the key genomic features associated with copy number heterogeneity was chromosomal instability as evidenced by WGD (p=0.003).

While multi-regional genomic analysis is an interesting research tool, it would be challenging to implement as a routine cancer diagnostic and is not part of the A151216 analysis. In contrast, single-sample assessment of WGD and TMB in resected NSCLC could be applied routinely in resection specimens, and could potentially be used as a risk stratification tool in future trials (while ITH has been shown to be prognostic and is of interest, it is difficult to calculate from the available data). WGD is a relatively common genomic event in NSCLC (~50%) and thus could be used to identify patients with high risk who might benefit more from an intensified adjuvant approach. Our overarching hypothesis is that such clinically practical genomic biomarkers can be derived in resected NSCLC and warrant further study as a poor risk feature in resected NSCLC.

For aim 3, Cohort C will be utilized and contains 2070 participants. 86% of Patients in Cohort C have been followed for at least 1 year with 39 % followed for 2 years or more. As this is an ongoing study, clinical follow-up is ongoing and patient status is continually being updated. To permit rigorous statistical analysis of evolving time-to-event endpoints like disease free survival (DFS) and overall survival (OS), the Alliance Statistics and Data Center will have sole access to clinical outcomes data. Genomic features derived by the GDAN will be shared with the Alliance SDC in order to study clinical outcomes in the A151216 population.

Statistical Considerations: The goal of this work is to first evaluate genomic factors that have been shown to be prognostic, and second to perform exploratory analyses in search of additional genomic prognostic factors. In both cases, the Cohort C exploratory and validation datasets defined below will be used for initial model estimation and then to perform independent replication, respectively. The clinical outcomes of interest are Disease Free Survival (DFS) and Overall Survival (OS). Overall survival time will be defined as the time from resection to death from any cause. Disease Free Survival will be defined as the time from resection to first relapse or death.
Design. A cohort study design will be employed utilizing the exploratory cohort described below. Baseline demographic and clinical variables will be compared between those going on to treatment trials versus those who do not to assess representativeness. If the non-treatment trial cohort differs systematically from those on treatment trials, weighting will be considered so that inferences can be applied to the entire screening population. Thus, the inferences will apply to the entire ALCHEMIST screening population.

Variables. *Clinical and demographic.* Known baseline clinical and demographic prognostic variables will be assessed: Stage (N stage, T stage), Tumor size, Age, Performance Status (PS) and smoking status (never, light, moderate, and heavy), Histology type, Gender (not typically significant), Receipt of adjuvant chemo (as a time-dependent variable), Surgical procedure

(lobectomy vs wedge vs other), weight loss over the 1 yr prior to diagnosis. *Genomic.* Primary genomic features of interest are WGD and TMB. Secondary genomic features of interest are predefined mSigDB gene sets defined previously herein. We will evaluate prognostic value of these genomic features over and above known prognostic factors. Targetable oncogenic driver mutations as listed in the NCCN guidelines will also be noted and considered. *Redundancy analysis.* Ignoring the OS and DFS dependent variables, a redundancy analysis will be used to eliminate redundant independent predictors from these clinical, demographic or genomic variables that are highly correlated with each other and thus not contributing independent information[7]. *Functional form.* While linear variables have been shown to capture the majority of an effect, it is plausible that there may be interactions with, for example, age, ethnicity, and histology [7, 8]. Thus, some non-linearities will be evaluated including interactions of age, ethnicity, and histology with clinical, demographic and genetic variables.

Model estimation. In the exploratory ALCHEMIST dataset (Cohort C) for both OS and DFS, penalized time to event regression (LASSO) will first be used together with cross validation to generate shrunken estimates of model parameters in order to minimize bias and maximize the likelihood of reproducibility[7-10]. The clinical and demographic variables will be assessed first, and variables with coefficients shrunk to zero will be removed. Second, the genomic variables, WGD and TMB, will be added to the retained (and forced to remain in the model) clinical and demographic variables both individually (to assess information added by each individually) and simultaneously (to assess whether they add unique information), and shrunken estimates generated for this set of variables. Third, the secondary gene set summaries will be considered in a similar fashion. Thus, the final discovery model will consist of the demographic and clinical variables retained from the first step, and the genomic variables retained from the second step, with possible gene set variables from the third step. Landmark predictions at 6 months, 2, 3, and 5 years will be generated. Modeling assumptions will be evaluated and appropriate actions taken to address violation of assumptions. Internal model validation and estimation of optimism will be performed using bootstrapping methods [7, 8].

Model assessment and replication. In the validation ALCHEMIST dataset (Cohort C), coefficients from the exploratory phase will be held fixed for the validation phase. Calibration will be assessed by comparing the observed and predicted values. Discrimination will be assessed using the c-statistic at different time points such as 6 months, 2, 3 and 5 years. All models will be tested for the appropriateness of the proportional hazards assumption, using both formal tests and visual examination of residuals over time.

Final model. Estimates of the final model parameters will be generated from the entire exploratory plus validation datasets combined for the maximum possible precision.

Power to detect various effects depends on the prevalence of each genomic factor and the number of levels. As a general guide, a sample size of 1102 patients provides 80% power to detect an effect reflected by a hazard ratio of 1.4 for a two-level factor with a prevalence of 30% vs. 70% if the DFS event rate in the control group is assumed to be 16% at 1 year. A hazard ratio of 2 means the event will occur twice as often at each time point given a one-unit increase in the predictor. Factors with higher prevalence would have greater power for an effect to be detected for the same sample size, conversely for those with lower prevalence. See tables below for sample sizes required to detect various HR using different prevalence and control rate estimates. Power will be higher for continuous variables than for dichotomous variables. WGD is a dichotomous variable with

~50% prevalence, while TMB is a continuous real number. One-year DFS was assumed to be between 0.12-0.20 for the control group.

Table 2: Total Sample size necessary to detect a given Hazard ratio with 80% power and 12% event rate in the control group

Prevalence (%)	Hazard Ratio					
	1.4	1.6	1.8	2.0	3.0	5.0
5		3421	2161	1535	575	241
10		1776	1118	792	295	123
20	1929	965	604	425		
30	1431	709	440	309		
40	1218	598	368	256		
50	1136	552	337	233		

Table 3: Total Sample size necessary to detect a given Hazard ratio with 80% power and 20% event rate in the control group

Prevalence (%)	Hazard Ratio					
	1.4	1.6	1.8	2.0	3.0	5.0
5		2154	1362	969	365	156
10		1120	706	501	188	81
20	1217	611	383	271		
30	905	451	281	198		
40	773	381	236	165		
50	723	354	217	151		

Management of Exploratory/Validation Cohorts

The exploratory and validation datasets will be updated every 6 months to add in additional follow-up from the original cohort. At the yearly dataset refresh time point, we will work with the sites to call in any overdue data. This dataset will then be frozen and used/reused as new hypotheses are proposed and approved by the A151216 Steering Committee. Data from the validation subsets will be partitioned based on specific hypotheses and used for individual analyses. The partition will be selected to minimize selection bias, e.g., based on patient date of enrollment or random selection. So for example, if 50 events (recurrences or deaths) for a given analysis are needed we will find the patient that reported the 50th event and include all patients enrolled up to the date that the patient reporting the 50th event was enrolled and make sure all expected visits have been accounted for. Power/sample size needs will be considered. For exploratory analyses the entire exploratory dataset will be provided. See **Figure 2** for the details.

Figure 2: Data flow Request Process

```
┌─────────────────────────────────────┐
│  Investigator submits analysis       │
│  request to Alchemist                │
│  Steering Committee                  │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│  Alchemist Steering Committee        │
│  reviews request and for             │
│  approved requests designates them   │
│  as exploratory or                   │
│  confirmatory                        │
└─────────────────────────────────────┘
        │                    │
        ▼                    ▼
  ┌──────────────┐     ┌──────────────┐
  │ Exploratory  │     │ Confirmatory │
  └──────────────┘     └──────────────┘
        │                    │
        ▼                    ▼
┌────────────────────┐  ┌──────────────────────────┐
│ Alliance SDC        │  │ Alliance SDC provides     │
│ provides full       │  │ subset of confirmatory    │
│ exploratory dataset │  │ dataset containing number │
│                     │  │ of patients needed based  │
│                     │  │ on hypotheses provided    │
└────────────────────┘  └──────────────────────────┘
        │                    │
        ▼                    ▼
┌────────────────────┐  ┌──────────────────────────┐
│ Analysis done by    │  │ Analysis done by          │
│ requesting          │  │ requesting investigator   │
│ investigator        │  │                           │
└────────────────────┘  └──────────────────────────┘
        │                    │
        ▼                    │
┌────────────────────┐       │
│ If signal found,    │       │
│ proposal sent to    │       │
│ Alchemist Steering  │       │
│ Committee for       │       │
│ request for         │       │
│ confirmatory dataset│       │
└────────────────────┘       │
        │                    │
        ▼                    ▼
┌──────────────────────────────────────┐
│  All manuscripts submitted to          │
│  Alchemist Steering                    │
│  Committee for review and approval     │
└──────────────────────────────────────┘
```

References

1. Bielski, C.M., et al., *Genome doubling shapes the evolution and prognosis of advanced cancers.* Nat Genet, 2018. **50**(8): p. 1189-1195.

2. Carter, S.L., et al., *Absolute quantification of somatic DNA alterations in human cancer.* Nat Biotechnol, 2012. **30**(5): p. 413-21.

3. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

4. Wang, C., et al., *PatternCNV: a versatile tool for detecting copy number changes from exome sequencing data.* Bioinformatics, 2014. **30**(18): p. 2678-80.

5. Campbell, J.D., et al., *Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas.* Nat Genet, 2016. **48**(6): p. 607-16.

6. Jamal-Hanjani, M., et al., *Tracking the Evolution of Non-Small-Cell Lung Cancer.* N Engl J Med, 2017. **376**(22): p. 2109-2121.

7. Harrell, F.E., Jr., *Regression Modeling Strategies With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Second ed. 2015: Springer.

8. Steyerberg, E., ed. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Statistics for Biology and Health, ed. M. Gail, et al. 2009, Springer.

9. Hastie, T., R. Tibshirani, and J.H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2001, New York, NY: Springer Science + Business Media LLC.

10. James, G., et al., *An Introduction to Statistical Learning: with Applications in R*. 2013, New York: Springer.