Implementation of Amplicon Parallel Sequencing Leads to Improvement of Diagnosis and Therapy of Lung Cancer Patients

Katharina König, PhD,†‡§ Martin Peifer,§‖ Jana Fassunke, PhD,*†‡Michaela A. Ihle, Dipl-Biol,*†‡Helen Künstlinger, Dpil-Biol,*†‡ Carina Heydt, MSc,*†‡ Katrin Stamm, PhD,¶ Frank Ueckeroth,*†‡ Claudia Vollbrecht, BSc,†‡ Marc Bos, MD,†‡ Masyar Gardizi, MD,†‡ Matthias Scheffler, MD,†‡ Lucia Nogova, MD,†‡ Frauke Leenders, MD,# Kerstin Albus, MSc,*†‡ Lydia Meder, MSc,*†‡ Kerstin Becker, PhD,** Alexandra Florin, MA,*†‡ Ursula Rommerscheidt-Fuss,*†‡ Janine Altmüller, PhD,**‖ Michael Kloth, MD,*†‡ Peter Nürnberg, Prof,** Thomas Henkel, PhD,¶ Sven-Ernö Bikár, PhD,†† Martin L. Sos, PhD,† William J. Geese, PhD,‡‡ Lewis Strauss,‡‡ Yon-Dschun Ko, MD,‡§ Ulrich Gerigk, MD,‡‖ Margarete Odenthal, PhD,*†‡ Thomas Zander, MD,†‡ Jürgen Wolf, Prof,†‡ Sabine Merkelbach-Bruse, MD,*†‡ Reinhard Buettner, MD,*†‡ and Lukas C. Heukamp, PhD, MD*†‡#****

Introduction: The Network Genomic Medicine Lung Cancer was set up to rapidly translate scientific advances into early clinical trials of targeted therapies in lung cancer performing molecular analyses of more than 3500 patients annually. Because sequential analysis of the relevant driver mutations on fixated samples is challenging in terms of workload, tissue availability, and cost, we established multiplex parallel sequencing in routine diagnostics. The aim was to analyze all therapeutically relevant mutations in lung cancer samples in a high-throughput fashion while significantly reducing turnaround time and amount of input DNA compared with conventional dideoxy sequencing of single polymerase chain reaction amplicons.

Methods: In this study, we demonstrate the feasibility of a 102 amplicon multiplex polymerase chain reaction followed by sequencing on an Illumina sequencer on formalin-fixed paraffin-embedded tissue in routine diagnostics. Analysis of a validation cohort of 180 samples showed this approach to require significantly less input material and to be more reliable, robust, and cost-effective than conventional dideoxy sequencing. Subsequently, 2657 lung cancer patients were analyzed.

Results: We observed that comprehensive biomarker testing provided novel information in addition to histological diagnosis and clinical staging. In 2657 consecutively analyzed lung cancer samples, we identified driver mutations at the expected prevalence. Furthermore we found potentially targetable *DDR2* mutations at a frequency of 3% in both adenocarcinomas and squamous cell carcinomas.

Conclusion: Overall, our data demonstrate the utility of systematic sequencing analysis in a clinical routine setting and highlight the dramatic impact of such an approach on the availability of therapeutic strategies for the targeted treatment of individual cancer patients.

Key Words: Lung cancer, Next-generation sequencing, Routine diagnostic, Formalin-fixed, Amplicon.

(*J Thorac Oncol.* 2015;10: 1049–1057)

Lung cancer is histologically divided into small-cell lung carcinomas and non–small-cell lung carcinomas (NSCLCs). NSCLC comprises approximately 85% of newly diagnosed lung cancers subclassified into adenocarcinoma (AD; ~50%), squamous cell carcinoma (SCC; ~30%), and large cell carcinoma (20%).[1] Recently, treatment paradigms shifted from those based on mere morphology to one that incorporates actionable genomic alterations.[2]

The discovery of actionable mutations in the *Epidermal Growth Factor Receptor* (*EGFR*) gene and rearrangements of the *Anaplastic Lymphoma Kinase* (*ALK*) that are primarily found in AD patients have led to a remarkable improvement in overall survival of genetically selected patients.[3,4] Also in SCC and small-cell lung carcinomas, several novel potential driver mutations are currently being evaluated as potential

*Institute of Pathology, University Hospital of Cologne, Cologne, Germany; †Center for Integrated Oncology Cologne/Bonn, University Hospital of Cologne, Cologne, Germany; ‡Lung Cancer Group Cologne, Department I for Internal Medicine, University Hospital of Cologne, Cologne, Germany; §Labor Dr. Quade und Kollegen GmbH, Cologne, Germany; ‖Institute of Human Genetics, University of Cologne, Cologne, Germany; ¶Targos Molecular Pathology GmbH, Kassel, Germany; #Department of Translational Genomics, University of Cologne, Cologne, Germany; **Cologne Center for Genomics, Cologne, Germany; ††GENterprise GENOMICS GmbH, Mainz, Germany; ‡‡Bristol-Myers Squibb R&D, Princeton, New Jersey; §§Evangelische Kliniken Johanniter, Bonn, Germany; ‖‖Malteser Krankenhaus, Lung Cancer Center, Bonn, Germany; ¶¶Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany; ##Institute for Hematopathology Hamburg, Hamburg, Germany; and ***New Oncology, Cologne, Germany.

König, Peifer, Fassunke, Buettner, and Heukamp contributed equally to this article.

Disclosure: The authors declare no conflict of interest.

Address for correspondence: Katharina König, Institute of Pathology, University Hospital Cologne, Kerpenerstraße 62, Köln 50937, Germany. E-mail: k_koenig@lab-quade.de

DOI: 10.1097/JTO.0000000000000570

ISSN: 1556-0864/15/1007-1049

actionable targets,[5,6] for example, amplification of *Fibroblast Growth Factor Receptor 1* (*FGFR1*).[7–9] However, the translation of these scientific advances is currently hampered by the limited availability of genomic information on tumor material in clinical routine diagnostics.

Due to whole genome sequencing efforts in lung cancer,[8,10–12] the number of newly identified potential driver mutations,[13] for example, *Fibroblast Growth Factor 3, RET*, and *ROS*,[14] in lung cancer has increased steadily.[15] The recently identified subgroups are small, comprising 0.5% to 3% of the lung cancer patients[16,17] compared with the most frequently activated oncogene mutation in the *V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog*[18] gene found in 15% to 30% of NSCLC patients.[19] Thus, routine molecular diagnostics for all actionable and potentially actionable driver mutations remains a growing challenge.

At present, mutation analysis of formalin-fixed paraffin-embedded (FFPE) tumor tissue is mainly based on single sequential polymerase chain reaction (PCR) followed by conventional dideoxy or Sanger sequencing (SS), which is both labor- and cost-intensive and limited by sensitivity and the amount of available DNA. Since we aimed to sequence the entire 3.2-kb coding region of *Discoidin domain-containing receptor 2 (DDR2)* to recruit patients into a clinical trial with Dasatinib entitled: "Trial of Dasatinib in Subjects With Advanced Cancers Harboring DDR2 Mutation or Inactivating B-RAF Mutation" (NCT01514864 clinicaltrials.gov), we opted to replace mutational analysis by SS with amplicon-based massive parallel sequencing.[20]

With next-generation sequencing (NGS) benchtop devices becoming available, we aimed to reduce the time required for comprehensive molecular diagnostics and minimize the amount of required FFPE-derived input DNA while at the same time increase the number of target regions analyzed.

MATERIALS AND METHODS

Sample Cohorts

Samples were analyzed as part of the standard diagnostic procedures in agreement with guidelines, with approval of the local ethics committee (Ref Number: 10–242) and diagnosed based on the 2004 World Health Organization classification of lung tumors[21] and the International Multidisciplinary Classification of Lung Adenocarcinoma.[22]

For DNA extraction, tumor areas were marked on a hematoxylin and eosin-stained slide by a senior pathologist. After deparaffination, tumor tissue was macrodissected from six unstained 10-µm sections and samples were lysed overnight with proteinase K at 37°C. Extraction was carried out automatically using a BioRobot M48 (Qiagen, Hilden, Germany) or a Maxwell 16 System (Promega, Madison, WI). Tumor content varied between 10% and 100%, with a mean value of 36.5% and a median of 30% (Fig. S3A, Supplemental Digital Content 2, http://links.lww.com/JTO/A840), and we found a significant correlation between tumor content and allelic frequencies of the identified variants (Fig. S3B, Supplemental Digital Content 2, http://links.lww.com/JTO/A840). Molecular diagnostics was performed at the accredited central molecular pathology laboratory using high-resolution melting prescreening and fragment length analysis as appropriate followed by SS.[21,23]

Ion AmpliSeq Custom DNA Panels and Library Preparation

DNA concentration was determined with the Qubit dsDNA HS Assay Kit (Life Technologies, Darmstadt, Germany) on the Qubit 2.0 Fluorometer. Up to 50 ng of FFPE DNA was amplified for 30 cycles with Ion AmpliSeq Custom DNA Panels (Life Technologies) covering 102 amplicons of 12 different genes (Table S1, Supplemental Digital Content 1, http://links.lww.com/JTO/A839) split in two primer pools with the 5× Ion AmpliSeq HiFi Master Mix, which is part of Ion AmpliSeq DNA Library Kit 2.0. After treatment with the FuPa reagent, PCR products of both primer pools from the same patient were pooled. PCR products were purified with 1.6× Agencourt AMPure XP beads (Beckman Coulter, Brea, CA) incubated for 5 minutes at room temperature, washed twice with 80% ethanol, and eluted in DNase/RNase free water. PCR products were incubated with NEXTflex DNA Adenylation Mix (Bioo Scientific, Austin, TX). NEXTflex DNA Barcode adapters (Bioo Scientific) were ligated to the fragments using Switch Solution and T4 DNA Ligase. Ligation products were cleaned up with 1.8× Agencourt AMPure XP beads incubated for 5 minutes at room temperature, washed twice with 80% ethanol, and eluted in DNase/RNase free water followed by size selection with 0.8× Agencourt AMPure XP beads incubated for 5 minutes at room temperature; the supernatant was collected and incubated with 0.2× Agencourt AMPure XP beads for 5 minutes at room temperature, washed twice with 80% ethanol, and eluted in DNase/RNase free water. The NEXTflex Primer Mix together with Platinum PCR SuperMix HiFi was used for the final PCR amplification with 10 cycles and a final clean-up with 1× Agencourt AMPure XP beads for 5 minutes at room temperature, washed twice with 80% ethanol, and eluted in DNase/RNase free water.

Sequencing Using the Illumina MiSeq Platform

Library concentration was determined with the Qubit dsDNA HS Assay Kit (Life Technologies) on the Qubit 2.0 Fluorometer, which varied from 5 to 25 ng/µl. Libraries were diluted to 10 nM and pooled in equimolar amounts. Pooled libraries (5 ng/µl) were spiked with 5% PhiX DNA (Illumina, San Diego, CA) and sequenced paired end with the "MiSeq reagent Kit V2 (300-cycles)" (Illumina). FastQ files generated by the MiSeq Reporter were used as data output.

Generation of Variant Lists and Alignment of Raw Sequencing Reads

Briefly, after removing adapter sequences, raw reads (FastQ files) were aligned against reference NCBI build 37 (human genome 19, hg19) and only against the chromosomal regions covered by our Custom Panel (CP) using Burrows Wheeler Alignment. Remaining unmapped reads were then realigned using BLAT[24,25] to detect longer insertions or deletions. After combining all alignments into a single BAM file, genomic variants were called using a modified version of our previously described method[8] and adapted for targeted