From the Journal *Nature*

nature.com > Journal home > Table of Contents

Letter

Nature **465**, 473-477 (27 May 2010) | doi:10.1038/nature09004; Received 7 December 2009; Accepted 10 March 2010

The mutation spectrum revealed by paired genome sequences from a lung cancer patient

William Lee[1], Zhaoshi Jiang[1], Jinfeng Liu[1], Peter M. Haverty[1], Yinghui Guan[2], Jeremy Stinson[2], Peng Yue[1], Yan Zhang[1], Krishna P. Pant[3], Deepali Bhatt[2], Connie Ha[2], Stephanie Johnson[4], Michael I. Kennemer[3], Sankar Mohan[5], Igor Nazarenko[3], Colin Watanabe[1], Andrew B. Sparks[3], David S. Shames[5], Robert Gentleman[1], Frederic J. de Sauvage[2], Howard Stern[4], Ajay Pandita[5], Dennis G. Ballinger[3], Radoje Drmanac[3], Zora Modrusan[2], Somasekar Seshagiri[2] & Zemin Zhang[1]

1. Department of Bioinformatics and Computational Biology, Genentech Inc., South San Francisco, California 94080, USA
2. Department of Molecular Biology, Genentech Inc., South San Francisco, California 94080, USA
3. Complete Genomics Inc., Mountain View, California 94043, USA
4. Department of Pathology, Genentech Inc., South San Francisco, California 94080, USA
5. Department of Oncology Diagnostics, Genentech Inc., South San Francisco, California 94080, USA

Correspondence to: Zemin Zhang[1] Email: zemin@gene.com

Abstract

Lung cancer is the leading cause of cancer-related mortality worldwide, with non-small-cell lung carcinomas in smokers being the predominant form of the disease[1, 2]. Although previous studies have identified important common somatic mutations in lung cancers, they have primarily focused on a limited set of genes and have thus provided a constrained view of the mutational spectrum[3, 4, 5, 6, 7, 8]. Recent cancer sequencing efforts have used next-generation sequencing technologies to provide a genome-wide view of mutations in leukaemia, breast cancer and cancer cell lines[9, 10, 11, 12, 13]. Here we present the complete sequences of a primary lung tumour (60× coverage) and adjacent normal tissue (46×). Comparing the two genomes, we identify a wide variety of somatic variations, including >50,000 high-confidence single nucleotide variants. We validated 530 somatic single nucleotide variants in this tumour, including one in the *KRAS* proto-oncogene and 391 others in coding regions, as well as 43 large-scale structural variations. These

constitute a large set of new somatic mutations and yield an estimated 17.7 per megabase genome-wide somatic mutation rate. Notably, we observe a distinct pattern of selection against mutations within expressed genes compared to non-expressed genes and in promoter regions up to 5 kilobases upstream of all protein-coding genes. Furthermore, we observe a higher rate of amino acid-changing mutations in kinase genes. We present a comprehensive view of somatic alterations in a single lung tumour, and provide the first evidence, to our knowledge, of distinct selective pressures present within the tumour environment.