Supplementary Materials For

Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity

Matthew T. Chang, Saurabh Asthana, Sizhi Paul Gao, Byron H. Lee, Jocelyn S. Chapman, Cyriac Kandoth, JianJiong Gao, Nicholas D. Socci, David B. Solit, Adam B. Olshen, Nikolaus Schultz, Barry S. Taylor

Supplementary Information

Supplementary Note 1

Among notable long-tail hotspots was E14K in Nucleoporin 93kDa (*NUP93*) (Fig. 4A). This highly expressed essential gene encodes a critical subunit of the nuclear pore complex. This hotspot was present in six breast cancers and one sample each of bladder, head and neck, hepatocellular, lung adenocarcinoma, and papillary thyroid cancers (Supplementary Fig. 8A, left). Among assessable breast cancers, these appear to arise in HER2-negative luminal tumors and is the fifth most commonly mutated gene in the 1303 breast cancers studied here (after hotspots in *PIK3CA*, *TP53*, *SF3B1* K700E, and *AKT1* E17K) (Supplementary Fig. 8A, right). Directly adjacent to E14K was a Q15* truncating hotspot, however, affected tumors expressed high levels of both the wildtype and mutant alleles. There was no detectable effect on gene expression of transcripts carrying a mutation predicted to trigger nonsense-mediated decay[1]. This is consistent with prior studies of loss-of-function alleles in human genomes[2], but contrary to the effect of such mutations in other cancer genes such as *TP53*[3] and even *CDKN2A* (Supplementary Fig. 7C).

Among other genes with two or more hotspots in the long tail, mutations in the MYC-associated factor X (MAX) were notable. *MYC* is an oncogene broadly implicated in the pathogenesis of multiple human cancers. While genomic amplification of *MYC* is common in many tumor types, *MYC* mutations are rare. We identified two *MYC* hotspots in this study (T58 and S146L), in one to three tumors each of head and neck cancers, lung adenocarcinomas, melanomas, lymphomas, neuroblastomas, colorectal cancers. However, MYC-mediated transformation through either activation or repression of MYC targets is dependent on its heterodimerization with MAX[4], which is an integral and constitutively expressed protein. It was notable, therefore, that we identified two *MAX* hotspots mutations (H28R and R60Q) in the helix-loop-helix (bHLHZ) DNA binding domain (Supplementary Fig. 8B). While recurrent germline *MAX* mutations have been reported in hereditary and sporadic pheochromocytoma and paragangliomas[5,6], these were truncating mutations at different residues compared to the somatic missense hotspots detected here (Supplementary Fig. 8B). The three dimensional structure of the MYC-MAX heterodimer revealed that the R60 and H28 interact with 5' CA and 3' G of the CACGTG E-box respectively (Supplementary Fig. 8C), indicating that the mutations target DNA binding of the complex rather than MYC dimerization. Notably, all four H28R mutations and 20% of the R60Q mutations arose in endometrial tumors spanning three of the four previously established subtypes, including one POLE-ultramutated, three MSI-H hypermutated, and two copy number-low endometrioid-like tumors. Moreover, we also identified in another copy number-low endometrial tumor a *MYC* H374R mutation that is homologous to *MAX* H28R (Supplementary Fig. 8C). The presence of these mutations in diverse cancer types and subtypes driven by very different underlying mutational processes indicates they are unlikely passengers due only to the mutational burden of the affected tumors. Finally, whereas the truncating germline mutations in *MAX* imply a tumor suppressor role, we found that *MAX* hotspots mutations were mutually exclusive with MYC mutations and genomic amplifications across affected tumor types (Supplementary Fig. 8D). This suggests that somatic *MAX* hotspots may be gain-of-function. However, due to the complexity of MYC function and the functional antagonism of MAX heterodimerization with MAD[7], functional validation is necessary.

Section References

1. Nagy, E. & Maquat, L.E. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* **23**, 198-9 (1998).
2. MacArthur, D.G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823-8 (2012).
3. Parikh, N. *et al.* Effects of TP53 mutational status on gene expression patterns across 10 human cancer types. *J Pathol* **232**, 522-33 (2014).
4. Amati, B. *et al.* Oncogenic activity of the c-Myc protein requires dimerization with Max. *Cell* **72**, 233-45 (1993).
5. Comino-Mendez, I. *et al.* Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nat Genet* **43**, 663-7 (2011).
6. Burnichon, N. *et al.* MAX mutations cause hereditary and sporadic pheochromocytoma and paraganglioma. *Clin Cancer Res* **18**, 2828-37 (2012).
7. Diolaiti, D., McFerrin, L., Carroll, P.A. & Eisenman, R.N. Functional interactions among members of the MAX and MLX transcriptional network during oncogenesis. *Biochim Biophys Acta* (2014).

References for Supplementary Table 1

1. Brennan, C. W. et al. The somatic genomic landscape of glioblastoma. Cell 155, 462–477 (2013).
2. Wu, G. et al. The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. Nature Genetics 46, 444–450 (2014)
3. Johnson, B. E. et al. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. Science 343, 189–193 (2014).
4. Pugh, T. J. et al. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature 488, 106–110 (2012).
5. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. Nature 500, 415–421 (2013).
6. Hovestadt, V. et al. Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. Acta Neuropathologica 125, 913–916 (2013).
7. Robinson, G. et al. Novel mutations target distinct subgroups of medulloblastoma. Nature 488, 43–48 (2012).
8. Pugh, T. J. et al. The genetic landscape of high-risk neuroblastoma. Nature Genetics 45, 279–284 (2013).
9. Molenaar, J. J. et al. Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. Nature 483, 589–593 (2012).
10. Sausen, M. et al. Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. Nature Genetics 45, 12–17 (2013).
11. Zhang, J. et al. Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. Nature Genetics 45, 602–612 (2013).
12. The Cancer Genome Atlas (TCGA). Comprehensive molecular portraits of human breast tumours. Nature 490, 61–70 (2012).
13. Stephens, P. J. et al. The landscape of cancer genes and mutational processes in breast cancer. Nature 486, 400–404 (2012).
14. Nik-Zainal, S. et al. Mutational processes molding the genomes of 21 breast cancers. Cell 149, 979–993 (2012).
15. Banerji, S. et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature 486, 405–409 (2012).
16. Nik-Zainal, S. et al. The life history of 21 breast cancers. Cell 149, 994–1007 (2012).
17. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 487, 330–337 (2012).
18. Seshagiri, S. et al. Recurrent R-spondin fusions in colon cancer. Nature 488, 660–664 (2012).
19. Li, M. et al. Whole-exome and targeted gene sequencing of gallbladder carcinoma identifies recurrent mutations in the ErbB pathway. Nature Genetics 46, 872–876 (2014).
20. Ahn, S.-M. et al. Genomic portrait of resectable hepatocellular carcinomas: Implications of RB1 and FGF19 aberrations for patient stratification. Hepatology 60, 1972–1982 (2014).
21. Fujimoto, A. et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. Nature Genetics 44, 760–764 (2012).
22. Biankin, A. V. et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 491, 399–405 (2012).

23. Wu, J. et al. Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways. Proc. Natl. Acad. Sci. U.S.A. 108, 21188–21193 (2011).
24. Jiao, Y. et al. DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. Science 331, 1199–1203 (2011).'
25. The Cancer Genome Atlas (TCGA). Comprehensive molecular characterization of gastric adenocarcinoma. Nature 513, 202–209 (2014).
26. Wang, K. et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. Nature Genetics 46, 573–582 (2014).
27. Zang, Z. J. et al. Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. Nature Genetics 44, 570–574 (2012).
28. The Cancer Genome Atlas (TCGA). Comprehensive molecular characterization of urothelial bladder carcinoma. Nature 507, 315–322 (2014).
29. Gui, Y. et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. Nature Genetics 43, 875–878 (2011).
30. Al-Ahmadie, H. et al. Synthetic lethality in ATM-deficient RAD50-mutant tumors underlies outlier response to cancer therapy. Cancer Discovery 4, 1014–1021 (2014).
31. Iyer, G. et al. Genome sequencing identifies a basis for everolimus sensitivity. Science 338, 221–221 (2012).
32. Davis, C. F. et al. The Somatic Genomic Landscape of Chromophobe Renal Cell Carcinoma. Cancer Cell 26, 319–330 (2014).
33. The Cancer Genome Atlas (TCGA). Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature 499, 43–49 (2013).
34. Guo, G. et al. Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. Nature Genetics 44, 17–19 (2012).
35. Peña-Llopis, S. et al. BAP1 loss defines a new class of renal cell carcinoma. Nature Genetics 44, 751–759 (2012).
36. Lee, R. S. et al. A remarkably simple genome underlies highly malignant pediatric rhabdoid cancers. J. Clin. Invest. 122, 2983–2988 (2012).
37. Grasso, C. S. et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature 487, 239–243 (2012).
38. Barbieri, C. E. et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nature Genetics 44, 685–689 (2012).
39. Baca, S. C. et al. Punctuated evolution of prostate cancer genomes. Cell 153, 666–677 (2013).
40. The Cancer Genome Atlas (TCGA). Integrated genomic analyses of ovarian carcinoma. Nature 474, 609–615 (2011).
41. Jones, S. et al. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. Science 330, 228–231 (2010).
42. The Cancer Genome Atlas (TCGA). Integrated genomic characterization of endometrial carcinoma. Nature 497, 67–73 (2013).
43. Ho, A. S. et al. The mutational landscape of adenoid cystic carcinoma. Nature Genetics 45, 791–798 (2013).
44. Dulak, A. M. et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. Nature Genetics 45, 478–486 (2013).
45. Song, Y. et al. Identification of genomic alterations in oesophageal squamous cell cancer. Nature 509, 91–95 (2014).

46. Stransky, N. et al. The mutational landscape of head and neck squamous cell carcinoma. Science 333, 1157–1160 (2011).
47. Agrawal, N. et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science 333, 1154–1157 (2011).
48. Lin, D.-C. et al. The genomic landscape of nasopharyngeal carcinoma. Nature Genetics 46, 866–871 (2014).
49. The Cancer Genome Atlas (TCGA). Integrated Genomic Characterization of Papillary Thyroid Carcinoma. Cell 159, 676–690 (2014).
50. De Keersmaecker, K. et al. Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. Nature Genetics 45, 186–190 (2013).
51. Holmfeldt, L. et al. The genomic landscape of hypodiploid acute lymphoblastic leukemia. Nature Genetics 45, 242–252 (2013).
52. Zhang, J. et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature 481, 157–163 (2012).
53. Landau, D. A. et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell 152, 714–726 (2013).
54. Quesada, V. et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nature Genetics 44, 47–52 (2012).
55. Puente, X. S. et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature 475, 101–105 (2011).
56. Morin, R. D. et al. Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. Blood 122, 1256–1265 (2013).
57. Love, C. et al. The genetic landscape of mutations in Burkitt lymphoma. Nature Genetics 44, 1321–1325 (2012).
58. Morin, R. D. et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature 476, 298–303 (2011).
59. The Cancer Genome Atlas (TCGA). Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. The New England journal of medicine 368, 2059–2074 (2013).
60. Ding, L. et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature 481, 506–510 (2012).
61. Yan, X.-J. et al. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nature Genetics 43, 309–315 (2011).
62. Beà, S. et al. Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. Proc. Natl. Acad. Sci. U.S.A. 110, 18250–18255 (2013).
63. Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D. & Nagata, Y. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature (2011).
64. Lohr, J. G. et al. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. Cancer Cell 25, 91–101 (2014).
65. Pickering, C. R. et al. Mutational Landscape of Aggressive Cutaneous Squamous Cell Carcinoma. Clinical Cancer Research 20, 6582–6592 (2014).
66. Hodis, E. et al. A landscape of driver mutations in melanoma. Cell 150, 251–263 (2012).
67. Krauthammer, M. et al. Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. Nature Genetics 44, 1006–1014 (2012).
68. Berger, M. F. et al. Melanoma genome sequencing reveals frequent PREX2 mutations. Nature 485, 502–506 (2012).
69. Stark, M. S. et al. Frequent somatic mutations in MAP3K5 and MAP3K9 in metastatic melanoma identified by exome sequencing. Nature Genetics 44, 165–169 (2012).

70. The Cancer Genome Atlas (TCGA). Comprehensive molecular profiling of lung adenocarcinoma. Nature 511, 543–550 (2014).
71. Imielinski, M. et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell 150, 1107–1120 (2012).
72. The Cancer Genome Atlas (TCGA). Comprehensive genomic characterization of squamous cell lung cancers. Nature 489, 519–525 (2012).
73. Peifer, M. et al. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nature Genetics 44, 1104–1110 (2012).
74. Rudin, C. M. et al. Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. Nature Genetics 44, 1111–1116 (2012).

Supplementary Tables

Supplementary Table 1: Summary of study cohort and tumor types analyzed

Supplied as external file

Supplementary Table 2: Summary of statistically significant hotspot mutations

Supplied as external file

Supplementary Table 3: Summary of presumptive false positive mutations

Supplied as external file

Supplementary Table 4: Statistical associations of mutant alleles and tumor types

Supplied as external file

Supplementary Table 5: GQ60GK and G60 mutations in Ras genes

Supplied as external file