



# Open Access to Large Scale Datasets Is Needed to Translate Knowledge of Cancer Heterogeneity into Better Patient Outcomes

## Citation

Beck, Andrew H. 2015. "Open Access to Large Scale Datasets Is Needed to Translate Knowledge of Cancer Heterogeneity into Better Patient Outcomes." PLoS Medicine 12 (2): e1001794. doi:10.1371/journal.pmed.1001794. <http://dx.doi.org/10.1371/journal.pmed.1001794>.

## Published Version

doi:10.1371/journal.pmed.1001794

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:14351212>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available. Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

EDITORIAL

# Open Access to Large Scale Datasets Is Needed to Translate Knowledge of Cancer Heterogeneity into Better Patient Outcomes

Andrew H. Beck\*

Cancer Research Institute, Beth Israel Deaconess Cancer Center, Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, United States of America

\* [abeck2@bidmc.harvard.edu](mailto:abeck2@bidmc.harvard.edu)



 OPEN ACCESS

**Citation:** Beck AH (2015) Open Access to Large Scale Datasets Is Needed to Translate Knowledge of Cancer Heterogeneity into Better Patient Outcomes. *PLoS Med* 12(2): e1001794. doi:10.1371/journal.pmed.1001794

**Published:** February 24, 2015

**Copyright:** © 2015 Andrew H. Beck. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** AHB was supported by funding from the Susan G. Komen for the Cure Foundation under Award Number CCR14302670 and the National Library Of Medicine of the National Institutes of Health under Award Number K22LM011931. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** AHB has read the journal's policy and has the following conflicts: AHB is on the Medical Advisory Board for Definiens and the Editorial Board of *PLOS Medicine*.

**Abbreviations:** DREAM, Dialogue for Reverse Engineering Assessments and Methods; HNSCC, head and neck squamous cell carcinoma; ICGC, International Cancer Genome Consortium; MATH, mutant-allele tumor heterogeneity; MEDICC, Minimum Event Distance for Intra-tumour Copy-

Cancer is a heterogeneous disease, which is comprised of a collection of diseases traditionally categorized by tissue type of origin. A distinct set of etiologic causes, treatments, and prognoses are associated with different cancers, and even within a given tissue type, cancer shows significant variability in molecular and clinical features across patients. This interpatient heterogeneity is a major rationale for large-scale research efforts (such as The Cancer Genome Atlas [TCGA] and the International Cancer Genome Consortium [ICGC]) to comprehensively profile the molecular landscape of patient cancer samples across all major cancers [1,2]. These efforts have been bolstered by the recent development of new genomic [3] and computational [4] technologies to enable increasingly detailed and comprehensive analyses of the molecular landscape of solid cancers. It is hoped that the comprehensive molecular characterization of large sets of cancer samples will lead to the identification of new therapeutic targets and the development of improved personalized therapies for cancer patients.

A major challenge in cancer therapy is the development of resistance to molecularly targeted therapies. Although targeted therapies may show initial benefit in the subset of patients carrying a targeted molecular alteration, most patients will nevertheless go on to develop resistance for most advanced solid cancers. Identifying and overcoming drug resistance represents one of the most significant challenges facing cancer researchers today [5]. It is increasingly recognized that cancer is not only a heterogeneous disease across patients but also a heterogeneous disease within individual patients, with different regions of a tumor showing different molecular features at the DNA, RNA, and protein levels [6–9]. This intratumoral molecular heterogeneity is hypothesized to be a major cause of drug resistance and treatment failure in cancer [10]. However, the clinical significance of intratumoral molecular heterogeneity is not yet well-defined, and assessment of intratumoral molecular heterogeneity is not currently used in clinical cancer medicine for assessing disease prognosis or guiding therapy. Two recent research articles published in *PLOS Medicine* show the potential clinical utility of measuring intratumoral genetic heterogeneity in clinical cancer samples.

In one, James Brenton, Florian Markowitz, and colleagues applied the Minimum Event Distance for Intra-tumour Copy-number Comparisons (MEDICC) algorithm they recently developed for phylogenetic quantification of intratumoral genetic heterogeneity from multiregion DNA copy number profiling data [11] to predict treatment resistance in high-grade serous ovarian cancer [12]. Their analysis suggests that multiregion tumor sampling, DNA copy number profiling, and quantification of intratumoral genetic heterogeneity with the MEDICC algorithm could be a useful approach for predicting patient survival in ovarian cancer, in which

number Comparisons; TCGA, The Cancer Genome Atlas.

**Provenance:** This is a Guest Editorial commissioned by the PLOS Medicine Editors; not externally peer reviewed.

higher levels of heterogeneity associated with decreased survival. This study provides data to support the long-standing hypothesis regarding treatment resistance and intratumoral genetic heterogeneity [10]. Although these results are promising, the developed approach requires sampling multiple distinct regions of tumor, which would be more expensive and complex than molecular profiling from a single tissue sample. It is not yet known how much tumor sampling will be required to adequately quantify intratumoral heterogeneity in the clinic or if measuring intratumoral heterogeneity from multiple tumor samples will outperform other molecular approaches (e.g., prognostic expression signatures [13,14]) for predicting response to therapy in ovarian cancer. These are important research questions that will need to be answered prior to clinical translation.

The second study comes from James Rocco and colleagues [15]. Previously, these investigators used a publicly available data set of whole exome sequencing data in head and neck squamous cell carcinoma (HNSCC) from Stransky et al. [16] to develop a simple quantitative measure of intratumoral heterogeneity (mutant-allele tumor heterogeneity [MATH]) and showed that MATH scores were higher in poor outcome classes of HNSCC [17]. In the current study, the authors used publicly available whole exome sequencing data provided by TCGA and showed that the MATH score is associated with prognosis in HNSCC and contributes additional prognostic information beyond that provided by traditional clinical and molecular features. Since the MATH score can be computed from whole exome sequencing data obtained from a single tumor sample (which is a data type that can be obtained from formalin-fixed, paraffin-embedded tumor tissue, as is routinely collected in pathology laboratories [18]), this approach may be more easily translated into clinical use, as compared with approaches requiring multiregion sampling and more complex computational algorithms for the assessment of intratumoral heterogeneity. Nonetheless, establishing the utility of the MATH score as an effective prognostic and/or predictive biomarker in HNSCC will require additional studies of the MATH score on well-controlled clinical cohorts comprised of homogeneously treated patients with tumors at specific head and neck anatomic locations. It is important to note that the development and application of MATH for assessing prognosis in HNSCC was based entirely on the analysis of publically available clinically annotated whole exome sequencing data, which demonstrates the value in making these data open to the community.

The continuing generation of high-quality, open-access Omics data sets from large populations of cancer patients will be critically important to enable the development of computational methods to translate knowledge of cancer heterogeneity into new diagnostics and improved clinical outcomes for cancer patients. As one step towards this goal, the DREAM (Dialogue for Reverse Engineering Assessments and Methods) consortium will use open innovation crowd sourcing to identify top-performing computational methods for inferring genetic heterogeneity from next-generation sequencing data provided by a large multi-institutional community of cancer genomics projects, including the ICGC and TCGA [19]. If successful, this open innovation competition may identify a set of best-in-class methods for measuring intratumoral genetic heterogeneity in cancer.

In parallel with these advances in computational methods for inferring intratumoral heterogeneity from genomics data, genomics technologies for measuring intratumoral heterogeneity at increasingly fine levels of granularity continue to improve. For example, recent advances in single-cell sequencing of DNA have provided detailed portraits of intratumoral genetic heterogeneity and clonal evolution in cancer [20,21], and recent advances in single-cell RNA sequencing [22], in situ RNA sequencing [23,24], and highly multiplexed next-generation immunohistochemistry [25–28] enable characterization of intratumoral heterogeneity in gene expression at a single cell level with subcellular resolution. Thus, there are now many options—both molecular and computational—for measuring and analyzing intratumoral molecular heterogeneity from clinical cancer samples.

Establishing the clinical utility of these new approaches for measuring intratumoral molecular heterogeneity will require applying these methods to large sets of archival tumor samples from randomized trials of cancer therapeutics [29] and high-quality prospective observational studies [30]. To maximize the value of the data that would be produced from such an undertaking, it is critical that infrastructure be created and supported to enable sharing of the Omics and clinical data with a large community of cancer researchers and data scientists. Ensuring open access to high-quality datasets will ensure that the largest possible community of researchers is able to address the most important problems in cancer medicine today. And in generating and sharing these data widely, we will massively increase our chances of effectively translating knowledge of intratumoral heterogeneity into meaningful advances for cancer patients.

## Author Contributions

Wrote the paper: AHB. ICMJE criteria for authorship read and met: AHB. Agree with manuscript results and conclusions: AHB.

## References

1. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, et al. (2010) International network of cancer genome projects. *Nature* 464: 993–998. doi: [10.1038/nature08987](https://doi.org/10.1038/nature08987) PMID: [20393554](https://pubmed.ncbi.nlm.nih.gov/20393554/)
2. Garraway LA, Lander ES (2013) Lessons from the cancer genome. *Cell* 153: 17–37. doi: [10.1016/j.cell.2013.03.002](https://doi.org/10.1016/j.cell.2013.03.002) PMID: [23540688](https://pubmed.ncbi.nlm.nih.gov/23540688/)
3. Meyerson M, Gabriel S, Getz G (2010) Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* 11: 685–696. doi: [10.1038/nrg2841](https://doi.org/10.1038/nrg2841) PMID: [20847746](https://pubmed.ncbi.nlm.nih.gov/20847746/)
4. Ding L, Wendl MC, McMichael JF, Raphael BJ (2014) Expanding the computational toolbox for mining cancer genomes. *Nat Rev Genet* 15: 556–570. doi: [10.1038/nrg3767](https://doi.org/10.1038/nrg3767) PMID: [25001846](https://pubmed.ncbi.nlm.nih.gov/25001846/)
5. Garraway LA, Jänne PA (2012) Circumventing cancer drug resistance in the era of personalized medicine. *Cancer Discov* 2: 214–226. doi: [10.1158/2159-8290.CD-12-0012](https://doi.org/10.1158/2159-8290.CD-12-0012) PMID: [22585993](https://pubmed.ncbi.nlm.nih.gov/22585993/)
6. Burrell RA, McGranahan N, Bartek J, Swanton C (2013) The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 501: 338–345. doi: [10.1038/nature12625](https://doi.org/10.1038/nature12625) PMID: [24048066](https://pubmed.ncbi.nlm.nih.gov/24048066/)
7. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, et al. (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 366: 883–892. doi: [10.1056/NEJMoa1113205](https://doi.org/10.1056/NEJMoa1113205) PMID: [22397650](https://pubmed.ncbi.nlm.nih.gov/22397650/)
8. Bashashati A, Ha G, Tone A, Ding J, Prentice LM, et al. (2013) Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J Pathol* 231: 21–34. doi: [10.1002/path.4230](https://doi.org/10.1002/path.4230) PMID: [23780408](https://pubmed.ncbi.nlm.nih.gov/23780408/)
9. De Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, et al. (2014) Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* (80: -) 346: 251–256.
10. Burrell RA, Swanton C (2014) Tumour heterogeneity and the evolution of polyclonal drug resistance. *Mol Oncol* 8: 1095–1111. doi: [10.1016/j.molonc.2014.06.005](https://doi.org/10.1016/j.molonc.2014.06.005) PMID: [25087573](https://pubmed.ncbi.nlm.nih.gov/25087573/)
11. Schwarz RF, Trinh A, Sipos B, Brenton JD, Goldman N, et al. (2014) Phylogenetic quantification of intratumour heterogeneity. *PLoS Comput Biol* 10: e1003535. doi: [10.1371/journal.pcbi.1003535](https://doi.org/10.1371/journal.pcbi.1003535) PMID: [24743184](https://pubmed.ncbi.nlm.nih.gov/24743184/)
12. Schwarz RF, Ng CKY, Cooke SL, Newman S, Temple J, Piskorz AM, et al. (2015) Spatial and Temporal Heterogeneity in High-Grade Serous Ovarian Cancer: A Phylogenetic Reconstruction. *PLoS Med* 12: e1001789. doi: [10.1371/journal.pmed.1001789](https://doi.org/10.1371/journal.pmed.1001789)
13. Waldron L, Haibe-Kains B, Culhane AC, Riester M, Ding J, et al. (2014) Comparative meta-analysis of prognostic gene signatures for late-stage ovarian cancer. *J Natl Cancer Inst* 106.
14. Riester M, Wei W, Waldron L, Culhane AC, Trippa L, et al. (2014) Risk prediction for late-stage ovarian cancer by meta-analysis of 1525 patient samples. *J Natl Cancer Inst* 106.
15. Mroz EA, Tward AM, Hammon RJ, Ren Y, Rocco JW (2015) Intra-tumor Genetic Heterogeneity and Mortality in Head and Neck Cancer: Analysis of Data from the Cancer Genome Atlas. *PLoS Med* 12: e1001786. doi: [10.1371/journal.pmed.1001786](https://doi.org/10.1371/journal.pmed.1001786)

16. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, et al. (2011) The mutational landscape of head and neck squamous cell carcinoma. *Science* 333: 1157–1160. doi: [10.1126/science.1208130](https://doi.org/10.1126/science.1208130) PMID: [21798893](https://pubmed.ncbi.nlm.nih.gov/21798893/)
17. Mroz EA, Rocco JW (2013) MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral Oncol* 49: 211–215. doi: [10.1016/j.oraloncology.2012.09.007](https://doi.org/10.1016/j.oraloncology.2012.09.007) PMID: [23079694](https://pubmed.ncbi.nlm.nih.gov/23079694/)
18. Van Allen EM, Wagle N, Stojanov P, Perrin DL, Cibulskis K, et al. (2014) Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat Med* 20: 682–688. doi: [10.1038/nm.3559](https://doi.org/10.1038/nm.3559) PMID: [24836576](https://pubmed.ncbi.nlm.nih.gov/24836576/)
19. Sage Bionetworks (2015) ICGC-TCGA DREAM Somatic Mutation Calling Challenge—Tumor Heterogeneity & Evolution. <https://www.synapse.org/#!/Synapse:syn2813581>. Accessed 15 January 2015.
20. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, et al. (2011) Tumour evolution inferred by single-cell sequencing. *Nature* 472: 90–94. doi: [10.1038/nature09807](https://doi.org/10.1038/nature09807) PMID: [21399628](https://pubmed.ncbi.nlm.nih.gov/21399628/)
21. Navin NE (2014) Cancer genomics: one cell at a time. *Genome Biol* 15: 452. doi: [10.1186/s13059-014-0452-9](https://doi.org/10.1186/s13059-014-0452-9) PMID: [25222669](https://pubmed.ncbi.nlm.nih.gov/25222669/)
22. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, et al. (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344: 1396–1401. doi: [10.1126/science.1254257](https://doi.org/10.1126/science.1254257) PMID: [24925914](https://pubmed.ncbi.nlm.nih.gov/24925914/)
23. Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, et al. (2014) Highly multiplexed subcellular RNA sequencing in situ. *Science* 343: 1360–1363. doi: [10.1126/science.1250212](https://doi.org/10.1126/science.1250212) PMID: [24578530](https://pubmed.ncbi.nlm.nih.gov/24578530/)
24. Ke R, Mignardi M, Pacureanu A, Svedlund J, Botling J, et al. (2013) In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* 10: 857–860. doi: [10.1038/nmeth.2563](https://doi.org/10.1038/nmeth.2563) PMID: [23852452](https://pubmed.ncbi.nlm.nih.gov/23852452/)
25. Rimm DL (2014) Next-gen immunohistochemistry. *Nat Methods* 11: 381–383. doi: [10.1038/nmeth.2896](https://doi.org/10.1038/nmeth.2896) PMID: [24681723](https://pubmed.ncbi.nlm.nih.gov/24681723/)
26. Angelo M, Bendall SC, Finck R, Hale MB, Hitzman C, et al. (2014) Multiplexed ion beam imaging of human breast tumors. *Nat Med* 20: 436–442. doi: [10.1038/nm.3488](https://doi.org/10.1038/nm.3488) PMID: [24584119](https://pubmed.ncbi.nlm.nih.gov/24584119/)
27. Giesen C, Wang HAO, Schapiro D, Zivanovic N, Jacobs A, et al. (2014) Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat Methods* 11: 417–422. doi: [10.1038/nmeth.2869](https://doi.org/10.1038/nmeth.2869) PMID: [24584193](https://pubmed.ncbi.nlm.nih.gov/24584193/)
28. Gerdes MJ, Sevinsky CJ, Sood A, Adak S, Bello MO, et al. (2013) Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc Natl Acad Sci U S A* 110: 11982–11987. doi: [10.1073/pnas.1300136110](https://doi.org/10.1073/pnas.1300136110) PMID: [23818604](https://pubmed.ncbi.nlm.nih.gov/23818604/)
29. Simon RM, Paik S, Hayes DF (2009) Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst* 101: 1446–1452. doi: [10.1093/jnci/djp335](https://doi.org/10.1093/jnci/djp335) PMID: [19815849](https://pubmed.ncbi.nlm.nih.gov/19815849/)
30. Ahern TP, Hankinson SE (2011) Re: Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst* 103: 1558–1559; author reply 1559–1560. doi: [10.1093/jnci/djr327](https://doi.org/10.1093/jnci/djr327) PMID: [21917608](https://pubmed.ncbi.nlm.nih.gov/21917608/)