# The Cancer Genome Atlas Head-Neck Squamous Cell Carcinoma Collection (TCGA-HNSC)

# Summary

**Redirection Notice**

This page will redirect to https://www.cancerimagingarchive.net/collection/tcga-hnsc/ in about 5 seconds.

The Cancer Genome Atlas Head-Neck Squamous Cell Carcinoma (TCGA-HNSC) data collection is part of a larger effort to build a research community focused on connecting cancer phenotypes to genotypes by providing clinical images matched to subjects from The Cancer Genome Atlas (TCGA). Clinical, genetic, and pathological data resides in the Genomic Data Commons (GDC) Data Portal while the radiological data is stored on The Cancer Imaging Archive (TCIA).

Matched TCGA patient identifiers allow researchers to explore the TCGA/TCIA databases for correlations between tissue genotype, radiological phenotype and patient outcomes. Tissues for TCGA were collected from many sites all over the world in order to reach their accrual targets, usually around 500 specimens per cancer type. For this reason the image data sets are also extremely heterogeneous in terms of scanner modalities, manufacturers and acquisition protocols. In most cases the images were acquired as part of routine care and not as part of a controlled research study or clinical trial.

## CIP TCGA Radiology Initiative

Imaging Source Site (ISS) Groups are being populated and governed by participants from institutions that have provided imaging data to the archive for a given cancer type. Modeled after TCGA analysis groups, ISS groups are given the opportunity to publish a marker paper for a given cancer type per the guidelines in the table above. This opportunity will generate increased participation in building these multi-institutional data sets as they become an open community resource. Learn more about the TCGA Head-Neck Phenotype Research Group.

## Acknowledgements

We would like to acknowledge the individuals and institutions that have provided data for this collection:

- Ontario Cancer Institute/University Health Network, Toronto, Ontario - Special thanks to **Fei-Fei Liu (MD), Kenneth Yip (Ph.D.), Katrina Rey-McIntyre (BSc.), Scott Bratman (MD, Ph.D.)** and **John Waldron (MD, MSc.).**
- University of Pittsburgh/UPMC, Pittsburgh, PA - Special thanks to **Margarita L Zuley**, **MD** and **Rose Jarosz**.
- University of North Carolina, Chapel Hill, NC - Special thanks to **J. Keith Smith**, **MD, Ph.D.** and **Shanah Kirk** from the Department of Radiology.
- MD Anderson Cancer Center, Houston TX - Special thanks to **Rivka Colen, MD**, **Kimberly Garcia**, and **Bowman Willams.**
- Vanderbilt University Medical Center, Nashville TN - Special thanks to **Dominique Delbeke, MD, Ph.D.** and **Michelle Pham**.

- Johns Hopkins University - Special thanks to **Paul Nagy, PhD, FSIIM** and **Gorkem Sevinc**
- University of Miami, Miller School of Medicine- Special thanks to **Jose M. Net, MD**
- Barretos Cancer Hospital, Barretos, São Paulo, Brazil – Special Thanks to **Fabiano Rubião Lucchesi, MD** and **Natália Del Angelo Aredes**.

**Data Access**

## Data Access

Some data in this collection contains images that could potentially be used to reconstruct a human face. To safeguard the privacy of participants, users must sign and submit a TCIA Restricted License Agreement to help@cancerimagingarchive.net before accessing the data.

| Data Type | Download all or Query /Filter | License |
|---|---|---|
| Images and Radiation Therapy Structures (DICOM, 130GB) | *Download Search*<br><br>(Download requires the NBIA Data Retriever) | TCIA Restricted |

Click the Versions tab for more info about data releases.

## Additional Resources for this Dataset

The NCI Cancer Research Data Commons (CRDC) provides access to additional data and a cloud-based data science infrastructure that connects data sets with analytics tools to allow users to share, integrate, analyze, and visualize cancer research data.

- Imaging Data Commons (IDC) (Imaging Data)
- Genomic Data Commons (GDC) (Genomic, Digitized Histopathology & Clinical Data)

## Third Party Analyses of this Dataset

TCIA encourages the community to publish your analyses of our datasets. Below is a list of such third party analyses published using this Collection:

- Crowds Cure Cancer: Data collected at the RSNA 2018 annual meeting (Crowds-Cure-2018)

**Detailed Description**

## Detailed Description

| Image Statistics | |
|---|---|
| Modalities | CT, MR, PT, RTDOSE, RTPLAN, RTSTRUCT |

| | |
|---|---|
| Number of Participants | 227 |
| Number of Studies | 479 |
| Number of Series | 2,561 |
| Number of Images | 270,376 |
| Images Size (GB) | 129.6 |

## Total Cases by Site

| Site | Patients |
|---|---|
| MDA (TSS code "-CV-") | 78 |
| UNC (TSS code "-BA-") | 22 |
| UPMC (TSS code "-CN-") | 5 |
| Vanderbilt (TSS code "-CR-") | 51 |
| JHU (TSS code "-BB-") | 15 |
| Miami (TSS code "-IQ-") | 5 |
| Barretos (TSS code "-UF-") | 16 |
| UHN (TSS code "-CQ-") | 35 |
| **Total** | **227** |

## Total Cases by Modality and Site

| Site | CT | MR | PT | **Grand Total** |
|---|---|---|---|---|
| JHU | 14 | 3 | 5 | **22** |
| MDA | 78 | 6 | 6 | **90** |
| UNC | 20 | 2 | 2 | **24** |
| UPMC | 5 | | 5 | **10** |
| Vanderbilt | 50 | 5 | 15 | **70** |
| Miami | 5 | | | **5** |
| Barretos | 16 | | | **16** |
| UHN | 23 | 12 | | **35** |
| **Grand Total** | **211** | **28** | **33** | **272** |

## GDC Data Portal - Clinical and Genomic Data

The Genomic Data Commons (GDC) Data Portal has extensive clinical and genomic data, which can be matched to the patient identifiers on the images here in TCIA. Below is a snapshot of clinical data extracted on 1/5/2016:

- TCGA-HNSC Clinical Data.zip

Case report forms for the clinical data collected by TCGA can be found on the Biospecimen Core Resource Clinical Data Forms linked below:

- Head and Neck Case Quality Control Form
- Head and Neck Enrollment Form
- Head and Neck Follow-Up Form

The contents of the clinical_patient file as of June 29, 2014, for cases whose imaging are archived on TCIA are here.

## A Note about TCIA and TCGA Subject Identifiers and Dates

**Subject Identifiers:** a subject with radiology images stored in TCIA is identified with a Patient ID that is identical to the Patient ID of the same subject with demographic, clinical, pathological, and/or genomic data stored in TCGA. For each TCGA case, the baseline TCGA imaging studies found on TCIA are pre-surgical.

**Dates:** TCIA and TCGA handle dates differently, and there are no immediate plans to reconcile:

- TCIA Dates: dates (be they birth dates, imaging study dates, etc.) in the Digital Imaging and Communications in Medicine (DICOM) headers of TCIA radiology images have been offset by a random number of days. The offset is a number of days between 3 and 10 years prior to the real date that is consistent for each TCIA image-submitting site and collection, but that varies among sites and among collections from the same site. Thus, the number of days between a subject's longitudinal imaging studies are accurately preserved when more than one study has been archived while still meeting HIPAA requirements.
- TCGA Dates: the patient demographic and clinical event dates are all the number of days from the index date, which is the actual date of pathologic diagnosis. So all the dates in the data are relative negative or positive integers, except for the "days_to_pathologic_diagnosis" value, which is 0 – the index date. The years of birth and diagnosis are maintained in the distributed clinical data file. The NCI retains a copy of the data with complete dates, but those data are not made available. With regard to other TCGA dates, if a date comes from a HIPAA "covered entity's" medical record, it is turned into the relative day count from the index date. Dates like the date TCGA received the specimen or when the TCGA case report form was filled out are not such covered dates, and they will appear as real dates (month, day, and year).

Citations & Data Usage Policy

## Citations & Data Usage Policy

Users must abide by the TCIA Data Usage Policy and Restrictions. Attribution should include references to the following citations:

**Data Citation**

Zuley, M. L., Jarosz, R., Kirk, S., Lee, Y., Colen, R., Garcia, K., Delbeke, D., Pham, M., Nagy, P., Sevinc, G., Goldsmith, M., Khan, S., Net, J. M., Lucchesi, F. R., & Aredes, N. D. (2016). **The Cancer Genome Atlas Head-Neck Squamous Cell Carcinoma Collection (TCGA-HNSC) (Version 6) [Data set]**. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.LXKQ47MS

**Acknowledgement**

"The results <published or shown> here are in whole or part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/."

**TCIA Citation**

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., & Prior, F. (2013). **The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository**. In Journal of Digital Imaging (Vol. 26, Issue 6, pp. 1045–1057). Springer Science and Business Media LLC. https://doi.org/10.1007/s10278-013-9622-7

## Other Publications Using This Data

TCIA maintains a list of publications which leverage our data.


1. Babier, A., Zhang, B., Mahmood, R., Moore, K. L., Purdie, T. G., McNiven, A. L., & Chan, T. C. Y. (2021). OpenKBP: The openaccess knowledgebased planning grand challenge and dataset. Medical Physics. doi:10.1002/mp.14845
2. Dalvit Carvalho da Silva, R. (2022). The Role of Transient Vibration of the Skull on Concussion. (Ph. D. ). University of Western Ontario, Retrieved from https://ir.lib.uwo.ca/etd/8399
3. Dalvit Carvalho da Silva, R., Jenkyn, T. R., & Carranza, V. A. (2021). Development of a Convolutional Neural Network Based Skull Segmentation in MRI Using Standard Tesselation Language Models. Journal of Personalized Medicine, 11(4), 310. doi:https://doi.org/10.3390/jpm11040310
4. Dalvit Carvalho da Silva, R., Jenkyn, T. R., & Carranza, V. A. (2021). Development of a Convolutional Neural Network Based Skull Segmentation in MRI Using Standard Tesselation Language Models. J Pers Med, 11(4). doi:10.3390/jpm11040310

5. Hou, Y. X., Ren, Z., Tao, Y. B., & Chen, W. (2021). Learning-based parameter prediction for quality control in three-dimensional medical image compression. []. Frontiers of Information Technology & Electronic Engineering, 22(9), 1169-1178. doi:10.1631/FITEE.2000234
6. Huang, C., Cintra, M., Brennan, K., Zhou, M., Colevas, A. D., Fischbein, N., . . . Gevaert, O. (2019). Development and validation of radiomic signatures of head and neck squamous cell carcinoma molecular features and subtypes. EBioMedicine, 45, 70-80. doi:10.1016/j.ebiom.2019.06.034
7. Kann, B. H., Hicks, D. F., Payabvash, S., Mahajan, A., Du, J., Gupta, V., . . . Aneja, S. (2020). Multi-Institutional Validation of Deep Learning for Pretreatment Identification of Extranodal Extension in Head and Neck Squamous Cell Carcinoma. J Clin Oncol, 38(12), 1304-1311. doi:10.1200/JCO.19.02031
8. Katsoulakis, E., Yu, Y., Apte, A. P., Leeman, J. E., Katabi, N., Morris, L., . . . Oh, J. H. (2020). Radiomic analysis identifies tumor subtypes associated with distinct molecular and microenvironmental factors in head and neck squamous cell carcinoma. Oral Oncology, 110, 104877. doi:https://doi.org/10.1016/j.oraloncology.2020.104877
9. Lv, W., Zhou, Z., Peng, J., Peng, L., Lin, G., Wu, H., . . . Lu, L. (2023). Functional-structural Sub-region Graph Convolutional Network (FSGCN): Application to the Prognosis of Head and Neck Cancer with PET/CT imaging. Computer Methods and Programs in Biomedicine. doi:https://doi.org/10.1016/j.cmpb.2023.107341
10. Mukherjee, P., Cintra, M., Huang, C., Zhou, M., Zhu, S., Colevas, A. D., . . . Gevaert, O. (2020). CT-based Radiomic Signatures for Predicting Histopathologic Features in Head and Neck Squamous Cell Carcinoma. Radiol Imaging Cancer, 2(3), e190039. doi:https://doi.org/10.1148/rycan.2020190039
11. Na, K. J., & Choi, H. (2018). Tumor Metabolic Features Identified by (18)F-FDG PET Correlate with Gene Networks of Immune Cell Microenvironment in Head and Neck Cancer. Journal of Nuclear Medicine, 59(1), 31-37. doi:10.2967/jnumed.117.194217
12. Nikolov, S., Blackwell, S., Zverovitch, A., Mendes, R., Livne, M., De Fauw, J., . . . Ronneberger, O. (2021). Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy: Deep Learning Algorithm Development and Validation Study. J Med Internet Res, 23(7), e26151. doi:10.2196/26151
13. Singh, A., Goyal, S., Rao, Y. J., & Loew, M. (2019). A Novel Imaging-Genomic Approach to Predict Outcomes of Radiation Therapy. (MS). George Washington University, https://scholarspace.library.gwu.edu/etd/kh04dq40j
14. Thomas, R., Schalck, E., Fourure, D., Bonnefoy, A., & Cervera-Marzal, I. (2021). 2Be3-Net: Combining 2D and 3D Convolutional Neural Networks for 3D PET Scans Predictions. Paper presented at the International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2021).
15. Vrtovec, T., Monik, D., Strojan, P., Pernuš, F., & Ibragimov, B. (2020). Autosegmentation of organs at risk for head and neck radiotherapy planning: from atlasbased to deep learning methods. Medical Physics, 47, e929-e950. doi: https://doi.org/10.1002/mp.14320

Wong, J., Fong, A., McVicar, N., Smith, S., Giambattista, J., Wells, D., . . . Alexander, A. (2019). Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. Radiother Oncol, 144, 152-158. doi:10.1016/j.radonc.2019.10.019

If you have a manuscript you'd like to add please contact TCIA's Helpdesk.

**Versions**

## Version 6 (Current): 2023/05/24

| Data Type | Download all or Query/Filter |
|---|---|
| Images and Radiation Therapy Structures (DICOM, 130GB) | *Download* *Search*<br><br>(Download requires the NBIA Data Retriever) |
| Tissue Slide Images (web) | *Search* |
| Clinical Data (TXT) | *Download* |
| Biomedical Data (TXT) | *Download* |
| Genomics (web) | *Search* |

Corrected modality for series: 1.3.6.1.4.1.14519.5.2.1.8421.4009.196228604563469888269110627731 .

## Version 5: 2020/05/29

| Data Type | Download all or Query/Filter |
|---|---|
| Images and Radiation Therapy Structures (DICOM, 130GB) | *Download*<br><br>(Download requires the NBIA Data Retriever) |
| Tissue Slide Images (web) | *Search* |
| Clinical Data (TXT) | *Download* |
| Biomedical Data (TXT) | *Download* |
| Genomics (web) | *Search* |

Updated clinical data link with latest spreadsheets from GDC. Added new biomedical spreadsheets from GDC.

## Version 4: Updated 2018/08/30

| Data Type | Download all or Query/Filter |
|---|---|

| | |
|---|---|
| Images (DICOM, 130GB) | *Download* (Download requires the NBIA Data Retriever) |
| Clinical Data (TXT) | *Download* |
| Genomics (web) | *Search* |

Added new DICOM images.

## Version 3: Updated 2017/03/30

| Data Type | Download all or Query/Filter |
|---|---|
| Images (DICOM, 108GB) | *Download*<br>(Download requires the NBIA Data Retriever) |
| Clinical Data (TXT) | *Download* |
| Genomics (web) | *Search* |

Added new DICOM images.

## Version 2: Updated 2016/01/05

| Data Type | Download all or Query/Filter |
|---|---|
| Images (DICOM, 88.1GB) | *Download*<br>(Download requires the NBIA Data Retriever) |
| Clinical Data (TXT) | *Download* |
| Genomics (web) | *Search* |

Extracted latest release of clinical data (TXT) from the GDC Data Portal.

## Version 1: Updated 2014/11/26

| Data Type | Download all or Query/Filter |
|---|---|
| Images (DICOM, 88.1GB) | *Download*<br>(Download requires the NBIA Data Retriever) |
| Clinical Data (TXT) | *Download* |
| Genomics (web) | *Search* |