Radiation Exposure Determination in a Secure, Cloud-based Online Environment

Ben C. Shirley[a], Eliseos J. Mucaki[b], Joan H.M. Knoll[a,c], Peter K. Rogan[a,b,d]*

[a]CytoGnomix Inc., London, Canada; [b]Department of Biochemistry, University of Western Ontario, London, Canada; [c]Department of Pathology and Laboratory Medicine, University of Western Ontario, London, Canada; [d]Department of Oncology, University of Western Ontario, London, Canada

*Correspondence:

Peter K. Rogan PhD
Departments of Biochemistry and Oncology
Schulich School of Medicine and Dentistry
University of Western Ontario
London ON N6A 2C1 Canada
E: progan@uwo.ca
P: 519-661-4255
https://orcid.org/0000-0003-2070-5254

Abstract

Background: In a major radiation incident, the speed of sample processing and interpretation of estimated exposures will be critical for triaging individuals. The Automated Dicentric Chromosome (DC) Identifier and Dose Estimator System (ADCI) selects and processes images to identify DCs and determines radiation dose without manual review. The goal of this study was to broaden accessibility and speed of this system with data parallelization while protecting data and software integrity.

Methods: ADCI_Online is a secure web-streaming platform that can be accessed worldwide from distributed local nodes. Data and software are separated until they are linked for estimation of radiation exposures. Performance is assessed with data from multiple biodosimetry laboratories.

Results: Dose estimates from ADCI_Online are identical to ADCI running on dedicated GPU-accelerated hardware. Metaphase image processing, automated image selection, calibration curve generation, and radiation dose estimation of a typical set of samples of unknown exposures were completed in <2 days. Parallelized processing and analyses using cloned software instances on different hardware configurations of samples at the scale of an intermediate-sized radiation accident (54,595 metaphase images) accelerated estimation of radiation doses to within clinically-relevant time frames.

Conclusions: The ADCI_Online streaming platform is intended for on-demand, standardized radiation research assessment, biodosimetry proficiency testing, inter-laboratory comparisons, and training. The platform has the capacity to handle analytic bottlenecks in intermediate to large radiation accidents or events.

Keywords: Radiation protection, Biodosimetry, Chromosome aberrations, Cytogenetics, Population Health, Digital Health Informatics, Cloud Computing, Streaming Services

Introduction

Interpretation of cytogenetic metaphase images and quantification of exposures have been labour intensive steps in cytogenetic biodosimetry, despite advances in computer-assisted dicentric chromosome (DC) recognition and workload sharing. In a moderate- to large-scale radiation exposure scenario, specimen volumes would likely overload available human resources and delay timely reporting of clinically significant overexposures. Once cell images are captured by a computerized-microscope system, the same computer typically performs image analysis, during which time the system is unavailable for accessing additional samples.

In the event that rapid examination of large numbers of samples is necessary, semi-automated analytic approaches requiring manual intervention (1) to eliminate unsuitable images or validate DCs flagged by the software can delay timely results. Outsourcing the analysis to the previously published Automated Dicentric Chromosome Identifier and Dose Estimator System (ADCI) software, which runs on a dedicated high-performance computer system, eliminates this bottleneck and significantly increases overall throughput (2). ADCI fully automates selection of suitable metaphase images, detection of DCs, and quantifies radiation exposure. ADCI first utilizes novel image processing segmentation techniques to locate and characterize features of chromosomes (3,4), followed by tandem machine learning to determine whether each chromosome is monocentric or dicentric (5,6). ADCI autonomously determines image selection criteria used to select optimal images for interpretation (3,4), constructs dose calibration curves,

and estimates whole or partial-body radiation dose in test samples with unknown exposures (7,8). The estimated time for semi-automated DC review was compared to unattended processing by ADCI in a scenario involving 1,000 potentially exposed samples (2). Only ADCI has the throughput to provide timely dose estimates for hundreds to thousands of patient samples in an intermediate to large-scale nuclear event, in contrast with all other approaches based on DC frequencies.

ADCI examined simulated population-scale exposures with datasets derived from ex vivo irradiated samples on a highly parallelized, multiprocessor supercomputer platform (2). Execution of ADCI on high-throughput hardware is one approach to increase overall image processing speed, however these systems are not widely available to the most likely end-users, biodosimetry laboratories, which are distributed worldwide. Here, an alternative strategy is implemented to analyze samples with an array of lower throughput systems which could feasibly achieve similar or faster processing times overall. Cloud computing offers on-demand availability of data storage and computing resources. It can be broadly defined as the outsourcing of information technology resources, such as computing hardware, storage, or software, to third parties which generally have vast pools of resources which can be accessed economically on an on-demand basis. These systems can adjust available resources and can therefore accommodate both intermittent need and bursts of resources when necessary. Cloud-based architectures have been utilized in a variety of healthcare-related systems including an imaging platform utilizing deep learning, cytogenetic and genomic analysis, digital pathology, and many other applications (9,10). When applied to metaphase image processing for biodosimetry, cloud computing allows

for the rapid allocation of computational resources proportional to the number of images to be analyzed, and deallocation of those resources when the need for them has subsided.

Increasingly, the advantages of cloud computing must be balanced against growing, unanticipated cybersecurity risks. Data security and regulatory considerations must be taken into account when designing these systems, a subject of much discussion and scrutiny (11). Internet security has become a focus of government and international organizations that have been targets of highly publicized security breaches (12–14). Data breaches have exposed patient health information; such systems must comply with appropriate regulations designed to safeguard protected health information such as the Health Insurance Portability and Accountability Act (HIPAA; United States), the Personal Health Information Protection Act (PIPEDA; Canada), General Data Protection Regulation (GDPR; European Union), and others. Failure to implement sufficient safeguards can lead to data breaches due to equipment theft, human error, hacking, ransomware attacks and misuse (15).

Biodosimetry laboratories are largely government chartered, which usually imposes strict regulations or required procedures for data processing and management. The ADCI_Online system proactively and securely enables cloud computing with ADCI using a zero-trust framework implemented by Amazon Web Services (AWS) (16). It can be accessed through AWS AppStream 2.0, a fully managed streaming service that provides secure access to desktop applications. Users upload locally stored metaphase images via an AWS Simple Storage Service (S3) bucket. These data are subsequently linked to ADCI within an interactive streaming session through a web browser. The experience is analogous to remotely accessing a Windows® system using desktop virtualization software. ADCI_Online remotely enables rapid and secure

estimation of radiation exposures of uploaded samples, obviating the requirement for a physical workstation in which ADCI has been installed. Separation of sensitive data from non-sensitive data and encryption of data classified as sensitive are key elements of cloud-based data protection (17). A similar framework could be utilized by others to make their own Windows®-based software and data scalable, available broadly, but securely protected from intrusion (see Data and Software Availability section).

Methods

Windows®-based ADCI has been ported to AWS AppStream 2.0, which is accessible worldwide through local AWS nodes. Operationally, ADCI_Online is indistinguishable from the version that runs on a dedicated, standalone computer. The online version is streamed with limited access to the file system. The cloud-based system is also configured by default with lower throughput hardware (2 vCPU, 3.75GB RAM) relative to the recommended specifications of a desktop computer running Windows®-based ADCI (Intel i7-6700HQ, 16GB RAM). As a result, processing of metaphase cell images is ~3 fold slower on a per sample basis when using the default cloud system configuration.

On-Demand Remote Access

For each new ADCI_Online streaming session, a software instance is cloned from a single base image, or “snapshot”, built using the AWS Image Builder tool. The snapshot is comprised of a Microsoft Windows® Server 2016 system with ADCI preinstalled. A snapshot can be used to clone streaming instances with distinct hardware configurations that are part of one or more “instance families”. In this manuscript, “stream.standard.medium” (standard), “stream.standard.large”, “stream.compute.large”, and the Graphics Processing Unit (GPU)-enabled “stream.graphics.g4dn.xlarge” (G4) AppStream 2.0 hardware configurations from the

“General Purpose”, “Compute Optimized”, and “Graphics G4dn” instance families were assessed for use with ADCI (Table 1). The same snapshot may be used to clone streaming instances from the “General Purpose” and “Compute Optimized” families, however it was necessary to create a second snapshot configured as described, but instead built on G4 hardware to allow “Graphics G4dn” instances to be cloned. By default, a streaming session boots using the standard hardware configuration, and although it has less computing power than a high-performance Microsoft Windows® system running ADCI, the cloud-based design of ADCI_Online allows for rapid expansion of resources as shown in Figure 1.

Security and Data Management

An AWS S3 bucket is utilized for storage of uploaded metaphase images and ADCI results. Each user is assigned to a unique folder of the S3 Bucket named according to the SHA-256 hash of their email address, which is therefore not displayed to other users as plain text; access to folders created for other users is disabled through the use of user-specific AWS Identity and Access Management (IAM) Policies. All access to the S3 bucket is disabled by default and each permissible action must be explicitly granted by the IAM policy, therefore if an IAM policy fails to attach to a user for any reason, they will not be granted access to S3 in any form. All persistent ADCI_Online user files are encrypted in transit to and from S3 storage (HTTPS protocol) and server-side encryption is applied to all files in the Bucket (AWS Key Management Service). Although sensitive information may exist in an unencrypted state during streaming while actively in use, AppStream 2.0 instances exist only for the duration of a password protected streaming session. When the session is completed, the streaming instance and its associated local file storage (Amazon Elastic Block Store) is deleted. This also means when a user creates a new streaming session, they are executing ADCI in a clean environment that

cannot be influenced by previous users (e.g., sensitive data or potentially malicious files cannot be carried forward to another user).

The cloud-based streaming instance executing ADCI is not internet accessible to third parties, cannot interact with the local file system or external devices, and files can only be transferred to/from the streaming instance by way of the S3 bucket. The user-specific folder in the S3 bucket is mounted to the cloud-based system only while an ADCI_Online streaming session is active. A Microsoft AppLocker policy disallows execution of all software or scripts except “whitelisted” software directly related to ADCI or AppStream 2.0 functionality. ADCI_Online runs exclusively on the cloud-based system and encrypted pixels are streamed to the user through a web browser. ADCI does not consume computing resources or store data on the user’s local computer system. Local keyboard and mouse commands are sent to the cloud-based system to control ADCI_Online. This prevents the inadvertent transfer of sensitive data to the local system, keeping it self-contained in S3. It is also a valuable layer of application-level security, as user interaction with ADCI is limited to software-curated keyboard and mouse inputs only.

Uploaded metaphase cell image data are solely owned by and may be removed by individual users. Program scripts automatically delete user data one month after uploading unless the user makes arrangements to maintain it. User data will not be examined by system administrators or stored at any other location unless it is explicitly pre-authorized. A program script examines and verifies uploaded file names and file headers of uploaded metaphase images

as either TIFF or JPG format, but no other information about uploaded files is accessed or recorded.

Interacting with ADCI_Online

Before an ADCI_Online streaming session begins, users can access a website we developed called the Data Storage and Retrieval web app (code available as a Zenodo archive: https://doi.org/10.5281/zenodo.5761745), which utilizes the AWS Javascript Software Development Kit (SDK) v2.977.0 to provide an interface to their user-specific S3 folder. The web app is accessed with a web browser (HTTPS protocol) and implements client-side Javascript allowing users to upload images directly to S3, meaning images are not first uploaded to our servers and subsequently transferred to S3. Users sign in with AWS Cognito credentials provided to them and upload metaphase cell images to their S3 folder. This mechanism also allows users to view the content of their S3 folder, delete metaphase images if desired, and download ADCI reports after they have been generated. A count of files successfully uploaded and files which failed to upload is displayed. In this way users can ensure all files were uploaded to the cloud storage successfully. AWS currently limits the connection between AppStream 2.0 instances and S3, allowing only the first 5,000 files in an S3 directory to be accessed by the streaming instance. This limit is not acceptable for ADCI_Online, as a single sample may contain more than 5,000 metaphase images. The Data Storage and Retrieval web app recognizes samples surpassing this file limit and overcomes it by automatically directing file uploads to subfolders within the newly created sample directory on S3. This process of splitting samples into multiple subfolders is not visible to the user and does not (and can not) modify local user files or directory structures.

After uploading of images is complete, a web link to initiate a streaming session and credentials to access ADCI are provided to the user. The “Stream View” session options are

either “Desktop” or “ADCI”. These options enable the web browser to display either a simulated Windows® Desktop (which can run ADCI) or standalone ADCI. Each new session automatically clones a streaming instance from the snapshot of the ADCI_Online system. Simultaneously, the user’s S3 folder is mounted to the streaming instance, allowing the user to access their uploaded metaphase images and save ADCI results as reports generated during the session. An overview of user interaction with ADCI_Online is presented in Figure 2.

The user-specific portion of the S3 bucket is mounted to the streaming instance and provides storage which precedes the activation of the streaming session and persists beyond the end of the session. Thus, a user may log out of a streaming session and retain all files when a new streaming session is activated. AWS requires a maximum streaming session duration to be specified and a session is automatically discontinued when this duration is reached. Since all user-specific files are retained between streaming sessions, session duration is generally an issue only when processing samples, as the result matrix for the sample being actively processed would be lost if a session is terminated. For this reason, both the duration of each streaming session and the disconnect timeout have been extended to 96 hours, the maximum duration permitted by AWS. If an unexpected disconnection occurs, the ADCI_Online session will continue running and wait for the user to reconnect until the disconnect timeout is reached. Once sample processing is initiated, a loss of internet connectivity will not cause ADCI_Online to shut down or discontinue processing samples within the disconnect timeout window. With large datasets requiring more than several hours to process, a new streaming session is recommended, as this will reinitialize the session duration timer, providing a full 96-hour window for image processing. A user should not log out of a streaming session during processing, but may close

their browser window and resume the session before the maximum duration or disconnect timeout is reached. When a sample is processed by the software, files specifying chromosome contours are generated for each image. These files indicate chromosome boundaries and classification (e.g., monocentric or DC) when viewing metaphase images in ADCI's microscope viewer. Because they are created rapidly during sample processing and are not properly synchronized with S3 due to their large numbers when written directly to persistent storage, ADCI_Online places them in the local directory of the streaming instance itself. A compressed set of these files is archived, compressed, and then copied to persistent storage when a sample is saved. The files are extracted and written to local storage when a sample is loaded. As described previously, the local storage of a streaming instance is deleted when a streaming session is complete. Additional details describing these interactions are provided in the ADCI manual (https://adciwiki.cytognomix.com) and instructions to access ADCI_Online are documented (https://radiation.cytognomix.com/adcionline).

Overview of ADCI Functionality

ADCI processes calibration and test samples of metaphase images originating from the same laboratory and programmatically identifies DCs (5). DC counts associated with the machine learning tuning parameter σ (a measure of sensitivity and specificity) are stored in an "adcisample" file generated during sample processing (18).

Image selection models consist of a set of filters and an optional image ranking method intended to exclude images with suboptimal chromosome morphology (3). The Image Selection Optimization Wizard examines a set of calibration samples and performs an exhaustive search of potential image selection model parameters, ranking generated models by p-value of Poisson fit,

curve fit residuals, or leave-one-out cross validation (7). Image selection model A_B selects the top 250 images scored by Group Bin Distance (which models chromosomes based on their normal length distributions). A_C selects the top 250 images ranked by combining Z-scores (based on the standard deviations from mean values of each filter for all images in the same sample) of Filters I–VI weighted according to [5,2,4,3,4,1], respectively. A_D applies standard deviation-thresholded image Filters I–VI. Automated178981 was generated using the Image Selection Optimization Wizard (leave-one-out evaluation method) and selects the top 300 images ranked by combining Z-scores of Filters I-VI which were weighted according to [5,4,1,4,5,3], respectively. C_B750 applies standard deviation-thresholded image Filters I–III and selects top 750 images ranked by Group Bin Distance. Image selection models are applied to processed sample files and images excluded by the model no longer contribute to DC or image counts when calculating DC frequency.

DC frequencies in processed calibration samples filtered with an image selection model were used to fit a calibration curve using the maximum likelihood method (19). The same image selection model was applied to processed test samples of unknown exposures (6). Estimated whole-body dose is estimated from the DC frequencies of test samples using the linear quadratic equation fit to the calibration curve. Estimated partial-body dose and fraction of blood irradiated is determined with the Contaminated Poisson method (20) after adjusting for false positives identified in unirradiated controls (8). ADCI can generate reports describing samples, calibration curves, image model optimization, and dose estimation on demand. These reports are exported to S3 storage where they may be downloaded to local systems. Screenshots and additional explanatory text providing an overview of the steps required to perform a typical analysis on

ADCI_Online, including the upload of metaphase images and creation of a new streaming session are indicated in Extended Data (https://doi.org/10.6084/m9.figshare.17129768).

Results

Image processing, image selection model optimization, calibration curve generation, and dose estimation capabilities of ADCI_Online were compared to output generated on a dedicated Windows®-based computer system and to previously published studies of the same samples. A set of calibration (10 samples, 50,497 images) and test (6 samples, 4,098 images) samples obtained from Health Canada (HC) were uploaded to cloud storage and processed with ADCI_Online (7). Additionally, a set of calibration and test samples obtained from Public Health England (PHE) were also uploaded (8). Calibration curves were generated, and whole-body exposures of HC samples were determined using the Dose Estimation Wizard for image section models A_B, A_C, A_D, and Automated178981 (7). Partial-body exposure levels were estimated for PHE samples, with each receiving the equivalent of 50% fractional exposures based on an associated dose calibration curve generated with image selection model C_B750 (8).

Validation of ADCI_Online

HC samples were uploaded and processed on the standard and G4 hardware configurations. Dicentric chromosome counts and frequencies at all σ values obtained with ADCI_Online were identical compared to those obtained through a standard MS Windows® version of ADCI installed on local computer hardware.

Exposures of uploaded HC and PHE samples (PHE_E: 4.0Gy actual partial-body dose, 50% irradiated fraction, PHE_F: 2.0Gy, 50%, PHE_G: 6.0Gy, 50%) were estimated with ADCI_Online. Dose estimates for homogeneously irradiated HC test samples (HS01, HS04,

HS05, HS07, HS08, HS10) were generated after application of 4 different image selection models (A_B, A_C, A_D, Automated178981) and matched previously published results (7). Estimates of partial-body dose and fraction of blood irradiated agreed with previously published results after application of the C_B750 image selection model (8). Note that slight variation in results are expected in every partial-body analysis in ADCI by a random sampling algorithm that removes false positive DCs in partial-body irradiated samples (8). Image Selection Model Optimization was performed on the full set of HC calibration samples both locally and on ADCI_Online. The top 10 image selection models found on both systems had identical model numbers, model scores, image exclusion filters, and image ranking methods.

Performance for Uploading and Processing Metaphase Images

A set of 16 HC samples containing 54,595 metaphase images (41.7GB on disk) was uploaded to cloud storage through our web app implemented using the AWS Javascript SDK. All uploaded HC samples were processed on a single standard instance and on a single G4 instance. Then, the samples were split into five groups of similar image count and examined simultaneously by five standard instances run in parallel. The times required to upload and process each sample are presented in Table 2. In contrast with the high performance computing version of ADCI (2), images from a single sample cannot be analyzed by multiple software instances, since this version was designed to process complete samples. Image counts analyzed by each of the five parallelized instances varied (Table 2, right column A-E) between 10,520 to 11,896. All samples completed processing with the standard AWS configuration in 46 hr 28 min and in 14 hr 24 min using the GPU-based G4 system. Processing times of the parallelized

instances ranged from 7 hr 43 min to 9 hr 20 min. Some HC calibration samples contain many more metaphase images than required by IAEA guidelines for cytogenetic biodosimetry (21).

The Data Storage and Retrieval web app, as configured, uploads files concurrently in 20 image batches. When all images in a batch have been uploaded, the process is repeated. Upload times can likely be improved by increasing batch size. The HC 0.25 Gy calibration sample was uploaded in 11.16 minutes with a batch size of 20. The same sample was uploaded again with a batch size of 30 in 7.55 minutes. We believe a batch size of 20 is a relatively conservative default value, and users connecting to ADCI_Online with high-performance computers may wish to increase it to accelerate upload speeds.

Two other AppStream 2.0 hardware configurations, “stream.compute.large" and “stream.standard.large”, were initially considered for further testing but were ultimately discarded in favour of the standard and G4 configurations. HC test samples were processed at a rate of 12.92 and 19.89 images/min on “stream.compute.large" and “stream.standard.large” systems, respectively. The “stream.compute.large" configuration performed poorly compared to standard, and the “stream.standard.large” configuration was only marginally faster than standard.

Speed of Overall Process

The estimated time required to process a typical set of calibration samples, determine optimal image selection models, process a typical set of test samples, estimate dose of the test samples, and other relevant operations is presented in Table 3. To estimate total processing time

for a set of samples, image processing rates of 19.58 and 97.53 images/min were respectively applied for a single standard instance and five standard instances executed in parallel. These are the rates at which ADCI_Online processed the set of HC samples. The image processing steps are the most time-intensive aspects in the analysis of a typical set of samples. Calibration sample image processing and image selection optimization steps are performed only once for each laboratory. Once a calibration curve and optimal image selection models have been established, they can be applied to any number of test samples. Steps listed under "Other Operations" require little to no processing time, therefore the time required to perform them is limited by the operator's knowledge of the system. On the other hand, the time required to process samples and generate optimal image selection models is limited by the processing power of the system executing ADCI. Overall, assuming only required steps are performed and a moderate level of operator experience with ADCI (2 hours to perform the steps listed under "Other Operations"), the estimated time required to perform all steps on a single standard instance is 17 hr 11 min (or 19 hr 30 min if the leave-one-out evaluation method is chosen when generating optimal image selection models). If five parallelized standard instances are used to process samples, the time required is reduced to 6 hr 32 min, or 8 hr 51 min using the leave-one-out evaluation method. Thus, interpretation of a typical cytogenetic biodosimetry dataset can be completed with ADCI_Online within 1-2 days. Image processing steps are the most time-consuming steps in an analysis, therefore relatively accurate estimates of the time required can be obtained by applying images/min analyzed by a certain hardware configuration to the number of metaphase images to be examined.

Discussion

ADCI_Online offers a secure, session-based service useful for radiation research, proficiency testing, inter-laboratory comparisons, and training. In a research context, the system could provide highly uniform, reproducible assessment in large studies of many individuals, for example those exposed to therapeutic radiation. Dose estimation can be carried out anywhere, as long as the internet connection is reliable.

Speed estimates for the major steps presented in Table 3 are conservative. For rapid processing of large numbers of samples or multiple calibration curves from different laboratories, image processing can be accelerated either by opting for more expensive high performance GPU hardware, cloning more ADCI_Online instances, thereby increasing the level of parallelization, or both. By default, ADCI_Online is executed on the AppStream 2.0 hardware configuration “stream.standard.medium”, which is suitable for typical applications. A range of significantly more powerful (and expensive) configurations are available. For example, compute optimized configurations for large datasets include “stream.compute.8xlarge” (32 vCPU, 60GiB RAM). However, it is likely that cloning multiple ADCI_Online sessions at lower throughput configurations and/or utilizing GPU-enabled instances would be a more efficient and cost-effective means of processing large numbers of samples, because the increased RAM present in higher throughput configurations has little to no effect on ADCI sample processing speed. Although ADCI utilizes multithreading, RAM usage is predominantly limited to storage of image files being actively examined and result matrices generated during sample processing. Any instance configuration can be cloned multiple times as the standard instance was while processing the HC sample set, thus multiple G4 processor instances could meet image processing demands for dose estimation of large numbers of samples. With robust network connectivity in

an emergency with numerous, potentially radiation exposed individuals, throughput and capacity for samples requiring simultaneous processing and dose evaluation can be expanded to seamlessly mitigate backlogs in sample interpretation.

Before a sample is processed in ADCI, users can enter information about the sample as freeform text. ADCI prevents the creation of sample or report files if they contain characters disallowed in Windows® file names. Depending on how users describe samples when they are entered into the ADCI system, “adcisample” files and reports generated by ADCI which contain results based on these samples might be construed as containing personal health information. For this reason, security considerations were designed into this system from the outset. Data to be analyzed and the software itself were deployed on different AWS platforms, which are only logically linked to one another during active streaming sessions when ADCI_Online is running on AppStream 2.0. Security of data was considered at every step of the development process, with images and results encrypted and stored in an S3 Bucket folder accessible only to the user that created them. The software is maintained in a dormant state that is inaccessible to the user unless explicitly authorized by the software provider. AppStream 2.0, which runs ADCI_Online, receives text and mouse commands from users, transmits encrypted pixels to the user’s output device and writes ADCI-generated, encrypted reports to the user’s S3 bucket. Users cannot directly upload files to or download files from the AppStream 2.0 system; therefore, all files must be transferred through the S3 bucket where the files are encrypted. Such a framework could be customized to conform with the needs of other cloud-based projects, regardless of the type of data that the software handles.

ADCI_Online uses cloud computing to increase throughput of biodosimetry through automated DC analysis and dose estimation of irradiated individuals. Biodose Tools (22), another web-based resource for estimation of radiation exposures, is not fully automated as it requires preprocessed counts of DCs for each sample. The health information that these software programs generate from potentially numerous individuals is both personalized and private. The design of ADCI_Online has emphasized a multidimensional approach to computer security and, to the extent possible, through local AWS nodes. These system properties suggest that streaming technology will be attractive for development of other high-throughput, medical software applications.

Fulfillment of compliance with national and regional health privacy standards (e.g., HIPAA, GPDR, PIPEDA) will be required for all aspects of handling of individual samples including inter-laboratory transfers (23), management of samples and data generated. Access to unencrypted data, which may contain protected health information in ADCI, occurs only during streaming. Persistently stored, sensitive data generated by ADCI is encrypted and may be deleted by the user at any time. Both AppStream 2.0 and S3 storage are HIPAA-eligible services, thus, future HIPAA compliance can be anticipated. S3 storage can be designated to be co-localized with users by AWS, ensuring that ADCI_Online meets data sovereignty requirements for national or regional privacy protection. Secure sample transfer, processing, and cloud-based data analyses between globally dispersed biodosimetry facilities will require careful implementation of protocols to adhere to these privacy regulations.

Interpretation of cytogenetic biodosimetry data can be performed by streaming ADCI_Online through a web browser, obviating the requirement for local high-throughput computing, software installation or configuration. ADCI_Online is particularly appropriate for short-term applications requiring rapid analysis of large numbers of samples. Image processing speed can be specifically tailored to meet individual user requirements, and parallelized cloud-based hardware can process samples at rates as fast or faster than dedicated high-performance computers. Persistent user data are encrypted at rest on cloud storage, in transit between S3 and ADCI_Online, and between S3 and the user's local system. When configured with an array of powerful GPU-based cloud systems, this system could efficiently perform radiation dose estimation of cytogenetic data at population-scale.

Acknowledgements

This project was supported by CytoGnomix Inc. and the University of Western Ontario.

Competing Interests

Ben C. Shirley is an employee and Peter K. Rogan and Joan H.M. Knoll are cofounders of CytoGnomix Inc. Eliseos J. Mucaki does not have any competing interests. The company has developed software which incorporates the methods presented in this study.

Data and Software Availability

Software

A Zenodo repository (https://doi.org/10.5281/zenodo.5761745) contains the Data Storage and Retrieval web app and a template AWS Identity and Access Management Policy described in this manuscript. This software is available under the terms of the GNU General Public License v3.0.

Extended Data

A PDF document on Figshare (https://doi.org/10.6084/m9.figshare.17129768) presents an overview of the steps required to perform a typical analysis on ADCI_Online. The estimated time required to perform each step is provided.

References

1. Schunck C, Johannes T, Varga D, Lörch T, Plesch A. New developments in automated cytogenetic imaging: unattended scoring of dicentric chromosomes, micronuclei, single cell gel electrophoresis, and fluorescence signals. Cytogenet Genome Res. 2004;104(1–4):383–9.

2. Rogan PK, Mucaki EJ, Shirley BC, Li Y, Wilkins RC, Norton F, et al. Automated cytogenetic biodosimetry at population-scale. Radiation. 2021 Jun;1(2):79–94.

3. Liu J, Li Y, Wilkins R, Flegal F, Knoll JHM, Rogan PK. Accurate cytogenetic biodosimetry through automated dicentric chromosome curation and metaphase cell selection. F1000Res. 2017 Aug 9;6:1396.

4. Shirley B, Li Y, Knoll JHM, Rogan PK. Expedited radiation biodosimetry by automated dicentric chromosome identification (ADCI) and dose estimation. J Vis Exp. 2017 Sep 4;127:e56245.

5. Rogan PK, Li Y, Wickramasinghe A, Subasinghe A, Caminsky N, Khan W, et al. Automating dicentric chromosome detection from cytogenetic biodosimetry data. Radiat Prot Dosim. 2014 Jun 1;159(1–4):95–104.

6. Li Y, Knoll JH, Wilkins RC, Flegal FN, Rogan PK. Automated discrimination of dicentric and monocentric chromosomes by machine learning-based image processing. Microsc Res Tech. 2016 May 1;79(5):393–402.

7. Li Y, Shirley BC, Wilkins RC, Norton F, Knoll JHM, Rogan PK. Radiation dose estimation by completely automated interpretation of the dicentric chromosome assay. Radiat Prot Dosim. 2019 Jan 9;186(1):42–7.

8. Shirley BC, Knoll JHM, Moquet J, Ainsbury E, Pham N-D, Norton F, et al. Estimating partial-body ionizing radiation exposure by automated cytogenetic biodosimetry. Int J Radiat Biol. 2020 Nov;96(11):1492–503.

9. Ali O, Shrestha A, Soar J, Wamba SF. Cloud computing-enabled healthcare opportunities, issues, and applications: a systematic review. Int J Inform Manage. 2018 Dec 1;43:146–58.

10. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. Nat Rev Clin Oncol. 2019 Nov;16(11):703–15.

11. Hathaliya JJ, Tanwar S. An exhaustive survey on security and privacy issues in Healthcare 4.0. Comput Commun. 2020 Mar 1;153:311–35.

12. Bronskill J. Chinese hackers infiltrated National Research Council computers [Internet]. CBC. 2014 [cited 2021 Dec 6]. Available from: https://www.cbc.ca/news/science/chinese-hackers-installed-malware-on-national-research-council-computers-1.2872385

13. Mills E. Hackers break into Large Hadron Collider computer [Internet]. CNET. 2021 [cited 2021 Dec 6]. Available from: https://www.cnet.com/tech/services-and-software/hackers-break-into-large-hadron-collider-computer/

14. Alonso-Zaldivar R. Hackers break into HealthCare.gov [Internet]. Yahoo. 2021 [cited 2021 Dec 6]. Available from: https://au.news.yahoo.com/hackers-break-into-healthcare-gov-24904466.html

15. Chernyshev M, Zeadally S, Baig Z. Healthcare data breaches: implications for digital forensic readiness. J Med Syst. 2018 Nov 28;43(1):7.

16. AWS Events. Delivering zero trust environments with Amazon AppStream 2.0 [Internet]. 2021 [cited 2021 Dec 7]. Available from: https://www.youtube.com/watch?v=6DgjQ78i0Jw

17. Chen D, Zhao H. Data security and privacy protection issues in cloud computing. In: 2012 International Conference on Computer Science and Electronics Engineering. Hangzhou, Zhejiang China; 2012. p. 647–51.

18. Rogan PK, Li Y, Wilkins RC, Flegal FN, Knoll JHM. Radiation dose estimation by automated cytogenetic biodosimetry. Radiat Prot Dosim. 2016 Dec 1;172(1–3):207–17.

19. Papworth DG. Curve fitting by maximum likelihood. Radiat Bot. 1975;15:127–40.

20. World Health Organization, International Atomic Energy Agency, editors. Handling of radiation accidents: proceedings of a symposium on the handling of radiation accidents. Vienna: International Atomic Energy Agency; 1969. 710 p. (Proceedings series).

21. International Atomic Energy Agency. Cytogenetic Dosimetry: Applications in Preparedness for and Response to Radiation Emergencies. Vienna: IAEA; 2011.

22. Hernández A, Endesfelder D, Einbeck J, Puig P, Benadjaoud A, Higueras M, et al. Biodose Tools: An R Shiny Application for Biological Dosimetry [Internet]. 2020. Available from: https://biodosetools-team.github.io/biodosetools/

23. Endesfelder D, Oestreicher U, Kulka U, Ainsbury EA, Moquet J, Barnard S, et al. RENEB/EURADOS field exercise 2019: robust dose estimation under outdoor conditions based on the dicentric chromosome assay. Int J Radiat Biol. 2021 Sep 2;97(9):1181–98.

Figures and Tables

Table 1. **AppStream 2.0 Instance Configurations Utilized to Test ADCI_Online**

Name	**Short Name**	**Instance Family**	**CPU**	**vCPUs**	**RAM (GiB)**	**GPU**
stream.standard.medium[a]	standard	General Purpose	Intel(R) Xeon(R) CPU E5-2686 v4 @ 2.30GHz	2	4	-
stream.standard.large	-	General Purpose	Intel(R) Xeon(R) CPU E5-2676 v3 @ 2.40GHz	2	8	-
stream.compute.large	-	Compute Optimized	Intel(R) Xeon(R) CPU E5-2666 v3 @ 2.90GHz	2	3.75	-
stream.graphics.g4dn.xlarge[a]	G4	Graphics G4dn	Intel(R) Xeon(R) Platinum 8259CL CPU @ 2.5GHz	4	16	NVIDIA® T4

[a] Configurations selected for HC sample set analysis

Table 2. **Time required to upload and process Health Canada samples**

Sample	Dose (Gy)	Images	Upload Time (min)	Processing time (min)			Standard x5: Instance designation[d]
				G4[a]	Standard[b]	Standard x5[c]	
Calibration	0	5484	5.06	83.82	269.92	237.23	D
Calibration	0.1	4225	3.85	63.40	200.46	187.63	E
Calibration	0.25	11896	11.16	189.26	612.45	559.77	A
Calibration	0.5	959	0.99	17.40	57.35	54.41	E
Calibration	0.74	10968	14.31	167.88	516.94	462.7	B
Calibration	0.98	5812	5.71	88.32	291.10	254.57	C
Calibration	1.46	1336	1.11	20.26	62.58	57.29	C
Calibration	1.96	2180	1.96	36.46	122.10	112.47	C
Calibration	2.92	3914	3.49	59.53	186.72	168.91	D
Calibration	3.9	3723	3.24	69.67	232.20	217.62	E
HS01	3.1	540	0.37	9.76	33.5	31.04	C
HS04	1.8	600	0.34	11.24	39.58	36.19	D
HS05	2.8	1136	0.82	17.53	59.42	53.79	E
HS07	3.4	477	0.31	9.29	31.68	29.06	E
HS08	2.3	637	0.40	9.43	33.79	30.37	D
HS10	1.4	708	0.52	10.71	38.38	34.91	C
Total		54595	53.64	863.96	2788.17	2527.96	559.77[e]

[a] Samples were processed sequentially on a single streaming instance with the “stream.graphics.g4dn.xlarge” hardware configuration. Table 1 contains system specifications for each hardware configuration.

[b] Samples were processed sequentially on a single streaming instance with the “stream.standard.medium” hardware configuration.

[c] Samples were distributed among five streaming instances executed in parallel, each with the “stream.standard.medium” hardware configuration.

[d] A letter designation was assigned to each of the five “stream.standard.medium” streaming instances executed in parallel (A-E). Each sample was processed in full on one of the five instances as indicated by the letter designation.

[e] Maximum processing time of the five “stream.standard.medium” streaming instances executed in parallel.

Table 3. **Estimated Time Required for a Typical ADCI_Online Session**

User Action	Time	
	Standard[c]	**Standard x5**[d]
Process calibration samples[a]		
7500 metaphase images in 7 calibration samples • 3 samples < 1 Gy (1500 images), 4 samples >= 1Gy (750 images)	6 h, 23 m	1hr, 17 m
Automated image selection model generation[a,b]		
Evaluation modes: Curve fit residuals or p-value of Poisson fit	1 h, 34 m	
Evaluation mode: Leave-one-out	3 h, 53 m	
Process test samples		
8500 metaphase images in 10 test samples • 7 homogeneously irradiated samples (700 images), 3 partially irradiated (1200 images)	7 h, 14 m	1 h, 27m
Other operations		
Optional - Review of processed samples in metaphase image viewer	variable	
Review of image selection models	30 – variable m	
Calibration curve generation	5 – 30 m	
Dose estimation	5 – 30 m	
Report generation and review	10 – 120 m	

[a] These steps must be performed only once for each set of calibration samples.
[b] Only one evaluation method is required for automated image selection model generation.
[c] Assumes image processing rate of 19.58 images/min
[d] Assumes image processing rate of 97.53 images/min

Figure 1. ADCI_Online leverages a cloud-based environment to expand system resources on-demand

When a streaming session is initiated, a new streaming instance is cloned from an existing snapshot of the ADCI_Online system. Thus, all instances are configured identically when first initiated. The strategies to expand throughput and rapidly increase processing in ADCI_Online include: (1) the system snapshot may be cloned as many times as necessary, providing an array of cloud-based systems available for use, and (2) expand computing resources on each streaming instance by leasing higher capacity hardware. Both options could also be implemented simultaneously.

Figure 2. User interactions with ADCI_Online.

(a) Before a new subscription period begins, the user signs into the Data Storage and Retrieval web app using AWS Cognito credentials provided by the system administrator to upload metaphase images to cloud storage (AWS S3). (b) The user then signs into AWS AppStream 2.0 in their web browser and requests a new ADCI_Online streaming session. Simultaneously, the user-specific S3 folder is mounted to the streaming instance, allowing the user to access their uploaded metaphase images, process samples and save results. (a) After sample processing, dose estimation and report generation with ADCI_Online, users again sign in to the Data Storage and Retrieval web app to access and download reports.

Fig. 1

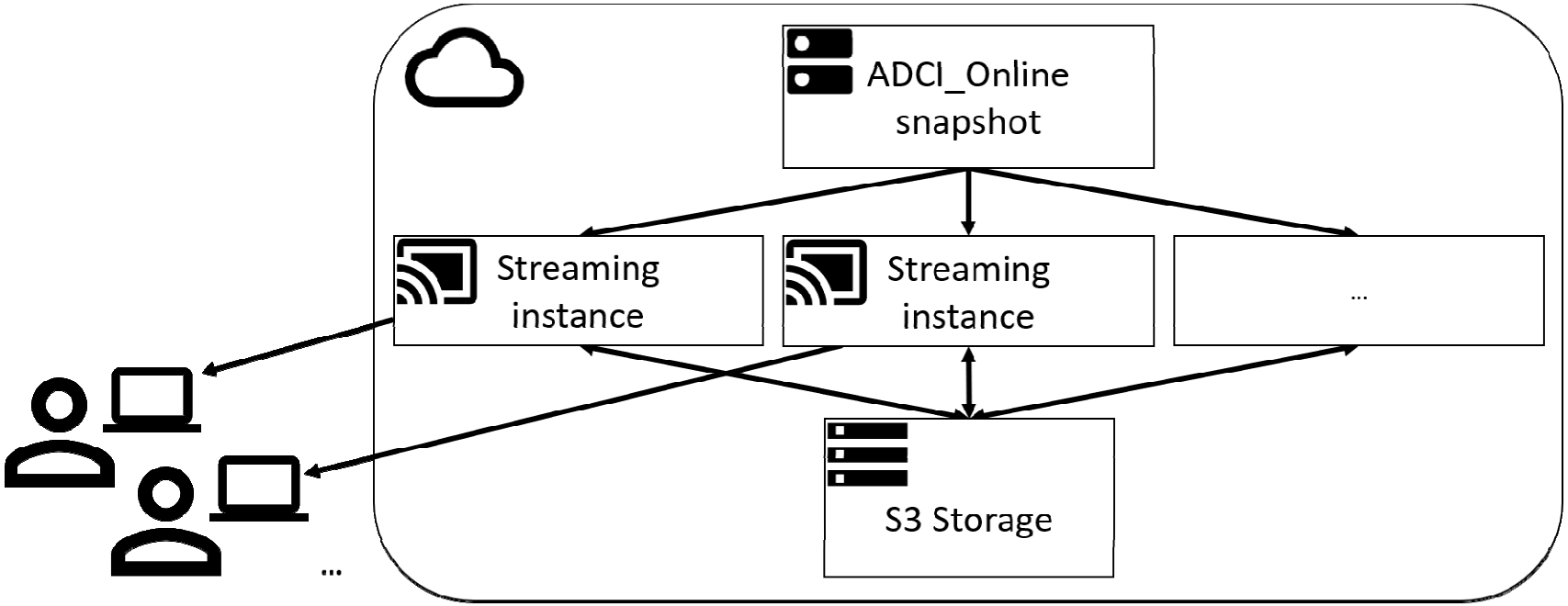
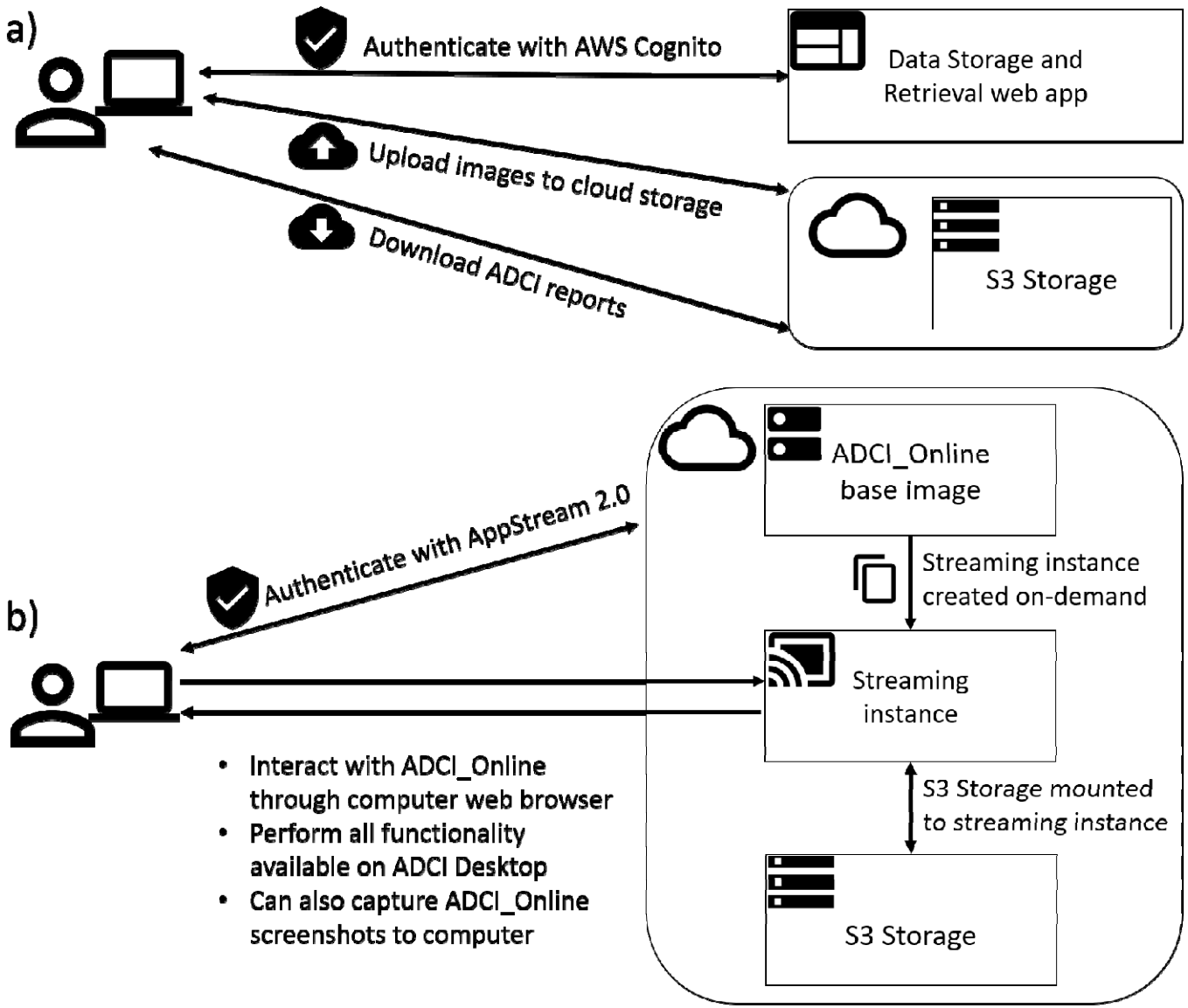

ADCI_Online snapshot
Streaming instance
Streaming instance
...
S3 Storage
...

Fig. 2.


a)
Authenticate with AWS Cognito
Data Storage and Retrieval web app
Upload images to cloud storage
Download ADCI reports
S3 Storage
b)
Authenticate with AppStream 2.0
ADCI_Online base image
Streaming instance created on-demand
Streaming instance
S3 Storage mounted to streaming instance
S3 Storage
• Interact with ADCI_Online through computer web browser
• Perform all functionality available on ADCI Desktop
• Can also capture ADCI_Online screenshots to computer