

Nutzungsordnung zum HPC-seq (Version vom 18.03.2019)

1. Einleitung

1.1. Bedeutung moderner Sequenzierverfahren an der MHH

Der Bedarf für moderne Sequenzierverfahren im Bereich von Grundlagen- und translationaler Forschung ist sowohl weltweit als auch an der Medizinischen Hochschule Hannover hoch und wächst stetig. An der MHH werden in verschiedenen Abteilungen DNA-Sequencer betrieben und entsprechend Sequenzdaten erhoben. Darüber hinaus ermöglicht die Research Core Unit Genomics (RCUG) als serviceorientierte Forschungseinrichtung der Gesamtheit aller Arbeitsgruppen der MHH Zugang zu moderner Sequenzier-technologie und sichert Unterstützung bei Projektplanung und Durchführung zu.

Mit wachsender Bedeutung entsprechender Sequenziersysteme und Diversifikation der Applikationen ergeben sich große Herausforderungen in den Bereichen Rechenleistung, Speicherkapazität, Langzeitdatenarchivierung und Datensicherheit. Für eine Vielzahl an der MHH etablierter Verbände und übergeordnete Strukturen stellt die sequenzbasierte Bioinformatik ein wichtiges Forschungsinstrumentarium dar (DZL, DZIF, SFB900, TRAINomics, TR209, Exzellenzcluster, CIIM,...).

Zur adäquaten Adressierung dieser Herausforderungen an der MHH bedarf es einer ausgereiften übergeordneten Konzeption, der Bündelung von Ressourcen und Expertisen sowie der engen kooperativen Zusammenarbeit der beteiligten Partner.

1.2. Konzept des HPC-seq

Der High Performance Computing Cluster zur (Nukleinsäure-)sequenzbasierten Bioinformatik (HPC-seq) stellt einen modular aufgebauten Verbund von Rechenressourcen dar. Durch die Zusammenlegung überwiegend drittmittelgeförderter Einzel-Ressourcen mehrerer Wissenschaftler und Einrichtungen der MHH konnte mit dem HPC-seq ein effizienter und zukunftsfähiger Rechencluster zur Verarbeitung von Primärdaten moderner Hochleistungs-Sequencer aufgebaut werden. Der HPC-seq stellt einen separaten, aber technisch kompatiblen Teil im IT-Konzept des ZIMts zum wissenschaftlichen Rechnen an der MHH dar.

Das Konzept des HPC-seq sieht eine zentral koordinierte Bereitstellung von Rechenleistung für sequenzbasierte Bioinformatik (Prozessierung und Auswertung von Sequenzdaten) für Forschungsvorhaben an der MHH vor.

Zur Berücksichtigung eines hinreichenden Maßes an Flexibilität und des nötigen Aufwuchspotentials wird der HPC-seq in Form eines modularen, erweiterbaren Systems betrieben. Dies ermöglicht eine skalierbare, effiziente und flexible Nutzung vorhandener Kapazitäten und kann den beteiligten Forschern bedarfsorientiert zur Verfügung gestellt werden. Darüber hinaus ergibt sich die

Möglichkeit der schrittweisen Erweiterung und Modernisierung der Hardware durch bestehende und neue Nutzer im Rahmen künftiger Forschungsanträge.

1.3. Initiative

Auf Initiative von RCUG-Lenkungsgruppe und ZIMt wurde im Jahre 2016 mit der Konzeption eines Drittmittel-finanzierten HPC-Clusters an der MHH begonnen. In der Folgezeit wurde das Projekt konkretisiert und schließlich in enger Zusammenarbeit von RCUG und ZIMt in Form eines HPC-Clusters zur bioinformatischen Sequenzanalyse (HPC-seq) etabliert und konfiguriert. Mehrere Arbeitsgruppen und Abteilungen der MHH, sowie an der MHH etablierte Forschungsverbände (Humangenetik, Pädiatrische Pneumologie (AG Tümmeler), Zahnärztliche Prothetik, RCUG, DZL, DZIF, SFB900) haben sich in dieser Etablierungsphase essentiell durch konzeptionelle Beiträge und Drittmittel-basierte Finanzierung einzelner Komponenten beteiligt.

1.4. Betrieb

Der Betrieb des HPC-seq wird durch wechselseitig verlässliche Zusammenarbeit des Zentrums für Informationsmanagement (ZIMt), der Research Core Unit Genomics (RCUG) und der HPC-seq Stakeholder (kurz Stakeholder) gewährleistet, die zu diesem Zweck eine Kooperationsvereinbarung geschlossen haben. Diese Erklärung regelt die jeweiligen Verantwortlichkeiten und Zuständigkeitsbereiche der Parteien und legt deren Befugnisse und Aufgaben verbindlich fest. Die Partner bekräftigen in dieser Vereinbarung darüber hinaus ihr gemeinsames Interesse am erfolgreichen Betrieb und der adäquaten, forschungsorientierten Weiterentwicklung des HPC-seq unter Berücksichtigung des Konzeptes zum Ausbau des Wissenschaftlichen Rechnens an der MHH.

2. Koordination und Steuerung

2.1. Wissenschaftlicher HPC-seq Koordinator

Als Wissenschaftlicher HPC-seq Koordinator (kurz Koordinator) wird Dr. Colin Davenport (RCUG) eingesetzt, der bis zu 20% seiner Arbeitszeit in diese Tätigkeit investiert. Insbesondere für alle anwendungsbezogenen Fragestellungen koordiniert Herr Dr. Davenport die Zusammenarbeit mit dem ZIMt sowie mit den Stakeholdern und sonstigen Nutzern und fungiert somit als primärer Ansprechpartner zur HPC-seq-Nutzung für wissenschaftliche Fragestellungen im Bereich Forschung und Lehre. Der Koordinator berichtet allen NutzerInnen auf den Nutzungsgruppentreffen (siehe unten) über die aktuelle Situation der Ressourcennutzung sowie über ggf. aufgetretene Probleme beim Anwendungsbetrieb.

Es liegt im Ermessen des Koordinators, Entscheidungen im Rahmen des Routinebetriebes in Absprache mit dem ZIMt entweder selbst zu treffen oder ein einstimmiges Votum der Stakeholder per E-Mail-Entscheid einzuholen. Sollte sich

ein solches Votum auf diese Weise nicht erzielen lassen, ist eine Abstimmung durch die Stakeholder auf dem nächsten Nutzungsgruppentreffen durchzuführen (nach angemessener zeitlicher Vorankündigung).

Dem Koordinator obliegt die Entscheidungsbefugnis, die HPC-seq Nutzung in einem vertretbaren Umfang auch „Nicht-Stakeholdern“ zu ermöglichen, solange i) alle Stakeholder entsprechend informiert sind, ii) die Rechte bei Fehlverhalten jederzeit entzogen werden können und iii) keine wesentliche Beeinträchtigung in der HPC-seq-Nutzung durch die Stakeholder resultiert. In jedem Einzelfall wird hierbei allerdings ein kurzer Antrag (ca. 1 DINA4-Seite) durch den Koordinator eingeholt, in dem der Anfragende die wesentlichen Details zur beabsichtigten HPC-seq Nutzung beschreibt.

Da die Verwendung des HPC-seq mit steigender Nutzung von immer fortschrittlicheren High-Output-Sequenzern kontinuierlich zunehmen wird, wurden Vorkehrungen zur Überwachung der HPC-seq-Nutzung getroffen, um eine effektive langfristige Quantifizierung der Ressourcennutzung zu ermöglichen. Zur kurzfristigen Überwachung der HPC-seq Nutzung ist eine Softwarelösung für Nutzer und Administratoren gleichermaßen verfügbar. Der Koordinator überprüft regelmäßig die jeweilige Ressourcennutzung.

2.2. Stakeholder und physikalische HPC-seq Zusammensetzung

Als Stakeholder im weiteren Sinne werden Verbände, Abteilungen, Arbeitsgruppen bzw. Antragsteller bezeichnet, die Komponenten zum HPC-seq beigesteuert (finanziert) haben. Hierbei gelten 10000,- Euro als Mindestbeitrag. Neben dem Erwerb von Hardware ist auch eine Beteiligung durch Anschaffung benötigter Softwarelizenzen oder über Personalmittel zur Stärkung der administrativen Aufgaben möglich.

Als Stakeholder im engeren Sinne sind jeweils ein Repräsentant per Verbund, Abteilung, Arbeitsgruppe, Antrag sowie sein Stellvertreter benannt, denen das jeweilige Stimmrecht obliegt.

Die jeweils aktuelle physikalische Zusammensetzung des HPC-seq aus einzelnen Teilkomponenten, sowie die Gesamtheit der Stakeholder (im weiteren und im engeren Sinne) wird in der Datei „HPC-seq Komponenten und Stakeholder_aktuell.xlsx“ ausgewiesen. Zum Zweck der notwendigen Aktualisierbarkeit, Gewährleistung von Transparenz und Nachverfolgbarkeit wird diese Datei auf dem MHH SharePoint-System bereitgestellt und von RCUG und Stakeholdern gemeinsam aktualisiert.

2.3. HPC-seq Nutzungsgruppe und Nutzungsgruppentreffen

Die HPC-seq Nutzungsgruppe (kurz Nutzungsgruppe) ist aus dem gesamten, den HPC-seq nutzenden Personenkreis zusammengesetzt.

Die Nutzungsgruppe trifft sich regelmäßig ca. alle 3 Monate, um über aktuelle Entwicklungen, Notwendigkeiten oder Probleme zu diskutieren und ggf.

erforderliche Entscheidungsprozesse vorzubereiten, welche durch Abstimmung der Stakeholder entschieden werden müssen.

Bei den Treffen der Nutzungsgruppe sollten alle Stakeholder durch mindestens einen stimmberechtigten Repräsentanten vertreten sein, insbesondere dann, wenn Abstimmungen auf der Agenda stehen.

Ausrichtung und Protokollierung der Nutzungsgruppentreffen wechseln reihum zwischen den Stakeholdern. Der Koordinator informiert jeweils nach dem abgehaltenen Treffen den Veranstalter des nächsten turnusmäßigen Treffens.

Das Protokoll der Nutzungsgruppentreffen wird allen Teilnehmern sowie allen Stakeholdern zugesandt und von der RCUG archiviert.

2.4. HPC-seq Entscheidungsprozesse

Sobald sich aus einem Nutzungsgruppentreffen die Notwendigkeit ergibt, eine wichtige übergeordnete Entscheidung zum HPC-seq treffen zu müssen, werden die entsprechenden Punkte im Vorfeld des nächsten Treffens auf die Agenda gesetzt und alle Stakeholder werden informiert, dass das nächste Nutzungsgruppentreffen wichtige Abstimmungsprozesse durch die Stakeholder beinhaltet.

Derartige Entscheidungen betreffen beispielsweise:

- die übergeordnete strategische Planung zur Erweiterung des HPC-seq
- die Konfliktlösung bei Problemen
- Engpässe bei der unmittelbaren oder mittelfristig antizipierten Ressourcennutzung
- die Beschlussfassung zur Aufnahme neuer Stakeholder
- erforderliche Änderungen in der Nutzungsordnung oder der Kooperationsvereinbarung mit dem ZIMt

Entscheidungen der Stakeholder bedürfen in Abstimmungen einer Zweidrittelmehrheit. Solche Entscheidungen dürfen allerdings in keinem Fall Aussagen zur Verwendung der HPC-Seq Komponenten aus den jeweiligen Forschungsanträgen zuwider laufen, über die die Komponenten angeschafft wurden, sonst gelten derartige Abstimmungsergebnisse als ungültig.

2.5. HPC-seq Sprecher

Die Stakeholder wählen unter sich für jeweils 2 Jahre einen HPC-seq Sprecher, der die Interessen und Belange der Stakeholder hochschulpolitisch vertritt. Des Weiteren überwacht der Sprecher gemeinsam mit der RCUG-Leitung und dem ZIMt die Einhaltung der Kooperationsvereinbarung und die wechselseitige Kompatibilität von HPC-seq-Konzeption und dem Konzept des wissenschaftlichen Rechnens an der MHH. Für die erste Zweijahresperiode (15.03.2019 bis

14.03.2021) wird durch die Stakeholder Prof. Dr. Thomas Illig zum HPC-seq Sprecher bestimmt.

3. Nutzungsbedingungen und Zugangsregelungen

3.1. Projektkategorien

Der HPC-seq ist in seiner derzeitigen Zusammensetzung und Konfiguration dezidiert zweckbestimmt und dem Einsatz im Rahmen von Forschungsvorhaben im Bereich der (Nukleinsäure-)sequenzbasierten Bioinformatik vorbehalten. Bei den Projektvorhaben werden folgende Kategorien unterschieden:

- 1: Forschungsvorhaben an nicht-menschlichem Untersuchungsmaterial oder an etablierten Zellkulturmodellen
- 2: Forschungsvorhaben an menschlichem Untersuchungsmaterial
- 3: Klinische Studien an menschlichem Untersuchungsmaterial

Projektvorhaben der Kategorien 1 und 2 können am HPC-seq durchgeführt werden. Für Vorhaben der Kategorie 3 sind insbesondere Fragen zur Studienfinanzierung durch Dritte (Wirtschaft) und der entsprechenden Nutzung von HPC-Komponenten (Verwendungsrichtlinien, finanzielle Beteiligung des Auftraggebers am HPC-seq) gesondert auszuarbeiten und den Stakeholdern vom PI des Projektes in schriftlicher Form vorzulegen. Es ist das erklärte Ziel, derartige Projektvorhaben auf dem HPC-seq zu ermöglichen.

3.2. Nutzer

Bei den Nutzern werden folgende Kategorien unterschieden:

- 1: Nutzer mit Finanzbeteiligung (Stakeholder im engeren und im weiteren Sinne)
- 2: Nutzer ohne Finanzbeteiligung

Die Nutzerverwaltung wird durch den Koordinator koordiniert. Eine entsprechend aktuelle Liste aller Nutzer wird vom Koordinator gepflegt und im SharePoint archiviert. Die Umsetzung der Nutzerverwaltung wird durch das ZIMt realisiert.

3.3. Garantierte Ressourcennutzung

Nutzern der Kategorie 1 steht das Recht zu, Rechenkapazität gemäß des eigenen Finanzbeitrags zu nutzen (im Folgenden als „Garantierte Ressourcennutzung“ bezeichnet).

Die Umsetzung der garantierten Ressourcennutzung wird vom Koordinator in Zusammenarbeit mit dem ZIMt in Form von Priorisierungen auf Nutzerebene geregelt (unterschiedlich priorisierte Warteschlangen).

Ebenfalls auf diesem Weg wird geregelt, in welcher Form Nutzern der Kategorie 2 Ressourcen bereitgestellt werden, ohne dass der priorisierte Anspruch der Stakeholder angetastet wird.

Über die jeweilige Strategie zur Priorisierung wird den Stakeholdern bei Bedarf berichtet. Die praktische Umsetzung wird zunächst so einfach wie möglich gehalten und folgt dem FAIR-Share Prinzip. Derzeit sind maximal drei verschiedene Priorisierungsstufen (gold, silber, bronze) geplant.

3.4. Eigentumsrechte

Aktuelle und künftige Stakeholder behalten die Eigentumsrechte an den jeweils beigesteuerten Komponenten. Außer dem hierdurch erworbenen Stimmrecht innerhalb der Steuerungsgruppe und der garantierten Ressourcennutzung verzichten sie aber auf jede darüber hinaus gehende eigenmächtige Entscheidung zur Verwendung der beigesteuerten Komponenten.

3.5. HPC-seq Zugang

Für den HPC-seq wurden drei Benutzerschnittstellen eingerichtet: Terminal, SMRT Link und Galaxy.

Erfahrene Linux-Benutzer können einen Terminal-Zugang verwenden. Da dieser sehr flexibel und programmierbar ist, ist das Terminal die universelle Schnittstelle zum HPC-seq. Die Web-Schnittstellen SMRT Link (für PacBio-Daten) und Galaxy (flexibel und erweiterbar für alle Arten von Bioinformatikdaten) ermöglichen unerfahreneren Benutzern den HPC-seq mit einer grafischen Oberfläche zu nutzen.

3.6. Konsequenzen bei Verstößen

Nutzer, die entweder gegen die in dieser Nutzungsordnung definierten Regelungen oder gegen mündliche oder schriftliche Anweisungen des Koordinators verstoßen, werden vom Koordinator in schriftlicher Form einmalig zur Einhaltung des Regelwerkes aufgefordert. Bei nochmaligem Verstoß erfolgt der Entzug der Zugangsrechte bei einem Nicht-Stakeholder endgültig, bei einem Stakeholder zunächst bis zur Entscheidung eines angemessenen Umgangs mit dem betreffenden Vorfall auf dem nächsten Nutzungsgruppentreffen.

4. Systemkonfiguration, Kernprozesse, Datensicherheit, Backup

Systemkonfiguration, Kernprozesse, Datensicherheit, Backup

Das Vorhalten und Aktualisieren von Dokumenten zum allgemeinen Betriebskonzept und zur Kernprozessbeschreibung des HPC-seq obliegt dem ZIMt.

Für alle Nutzer gelten die allgemeinen Bestimmungen des MHH Sicherheitskonzeptes und des MHH Datenschutzes. Benutzer sind verantwortlich für eigene Datensicherungen und Einhaltung entsprechender Vorhaltefristen. In der Regel werden nur scratch Partitionen ohne Backup eingesetzt (siehe http://hpc-web1.mh-hannover.local/doku.php?id=data_backup)

Das ZIMt bietet Dienste zur Sicherung und Archivierung von Daten an.

5. Erweiterung des HPC-seq

Künftige Beschaffungsmaßnahmen

Je nach aktueller Auslastung und Bedarf ist es das ausdrückliche Ziel, den HPC-seq fortwährend anzupassen und zweckdienlich zu erweitern. Ziel ist hierbei, dass Antragstellern, die für ihre NGS-Vorhaben relevante Rechenressourcen benötigen, frühzeitig die Vorzüge der Integration ihres Projektes in den HPC-seq verdeutlicht werden. So sollen Beantragungen von „IT-Insellösungen“ möglichst vermieden werden. Durch dieses Vorgehen können zum gegebenen Zeitpunkt gezielt notwendige Komponenten über den entsprechenden Neuantrag des jeweiligen neuen Stakeholders akquiriert und implementiert werden. Die hier vorliegende Nutzungsordnung ermöglicht dabei Antragstellern und potentiellen Drittmittelgebern sich frühzeitig über die genaue Verwendung der beantragten Komponenten und über das HPC-seq Konzept zu informieren.