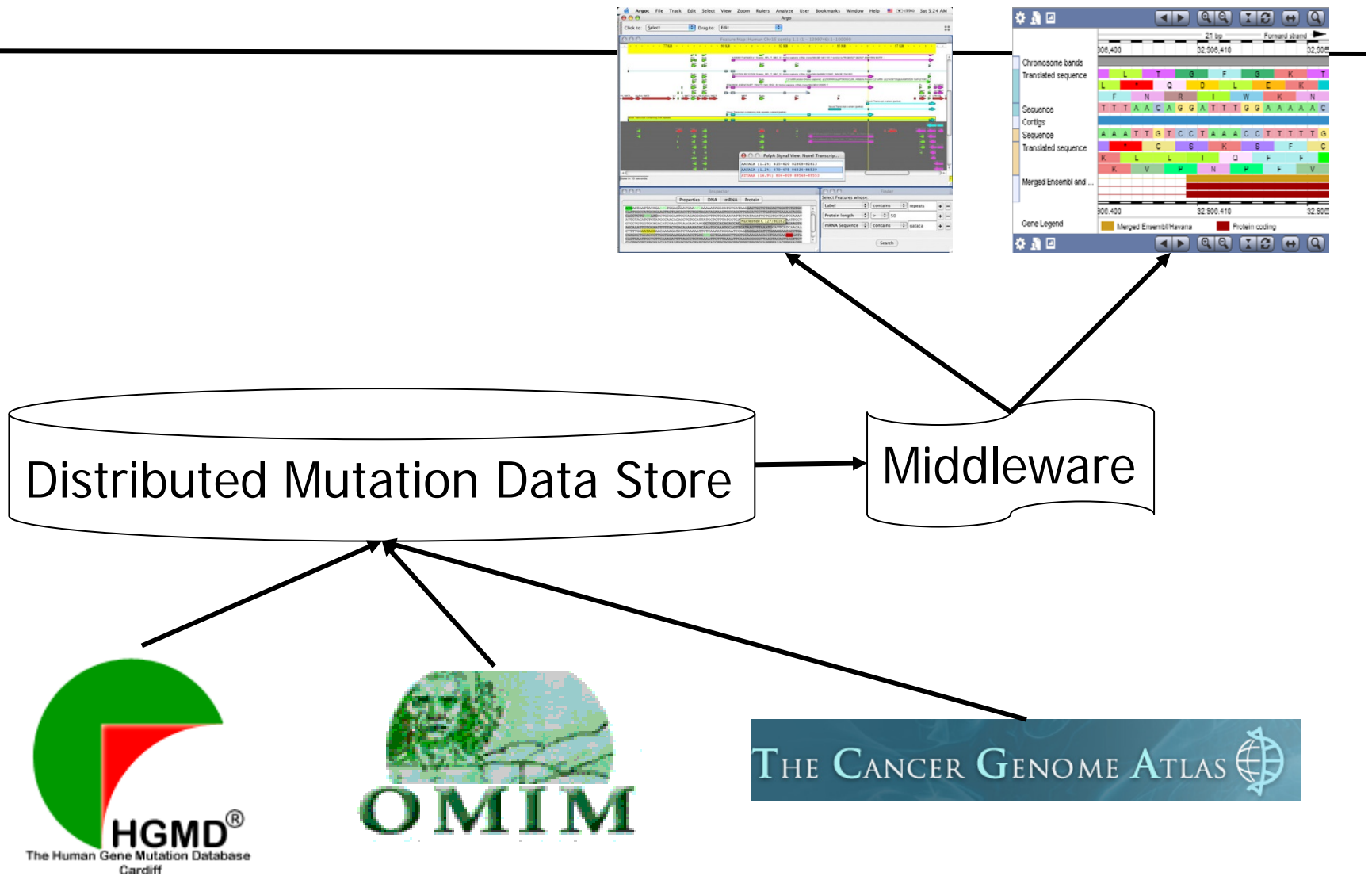




Semesterprojekt Anforderungen

Ulf Leser, Marc Bux, Stefan Sprenger

Genombrowser



Datenquellen

- Mindestens
 - HGMD
 - dbSNP
 - 1000 Genomes Projekt
 - The Cancer Genome Atlas: Lung and Colorectal cancer
- Wir wollen wissen
 - Metadaten: Datenquelle, Zeitpunkt Download
 - Position der Mutation (ggf. Abschnitt)
 - Mutierte Basensequenz (auch Referenz muss bekannt sein)
 - Häufigkeit
 - Quell-spezifische Sample-Attribute (Krankheit, Geschlecht, ...)

Anfragen

- **Intervall**
 - Gegeben Chromosom sowie linke und rechte Koordinate
 - Alle Mutationen in diesem Bereich mit Attributen
 - Zahl Mutationen im Bereich
 - Intervalle nie größer als 10MB
- **... oder Gen**
 - Gegeben der Name eines humanen Gens
 - Ggf. Nachfrage, wenn Name nicht eindeutig (Homonym)
 - Intervallgrenzen werden automatisch bestimmt (RefSeq)
- **Filter**
 - Einschränkung auf Quelle
 - Einschränkung auf relativer Häufigkeit
 - Bedingungen auf den quellspezifischen Attributen (nur equal)

Visualisierung

- Das humane Genom, Referenzsequenz
- Alle Mutationen im angezeigten Bereich
- Möglichkeit zum Zoomen (rein / raus – max 10MB)
- Bei großen Bereichen – summary view, die die Zahl der Mutationen in kleineren Bereichen andeuten
- Eine Lane pro Datenquelle

Non-Functional

- Erweiterbar, insb. für **neue Datenquellen**
 - Klares und einfaches „howto“ muss vorliegen
- Einfach zu installieren
- Konfigurierbar (Datenbankverbindung, Eigenschaften der Visualisierung, etc.)
- **Schnell: Alle Queries unter 1 sec (bis Anzeige)**
- Firefox, aktuelle Version, alle OS
- Integrierbar in bestehende Webserver (Standard-API's verwenden)

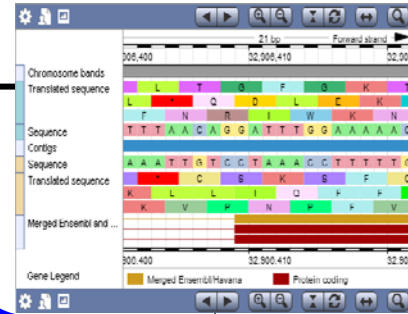
Bis Montag, 26.10.15, 9.00 Uhr

- Angegebene Literatur lesen!
- Einen Bericht organisieren
- Projektstruktur (Teilprojekte, Projektleitung)
 - Zuordnung Studierende – Teilprojekte
 - Grobe Charakterisierung der Aufgaben pro Teilprojekt
- Informelle Beschreibung von 3 Use Cases
 - Welche biomedizinische Aufgabe kann man mit dem Tool wie lösen? Was muss man dazu tun, was bekommt man gezeigt?
 - Nur „sichtbare“ Teile – weniger, was intern passiert (stört aber auch nicht)
 - Wir sehen uns gerne Entwürfe an
- Geplante Projektmanagementtools
 - Versionsverwaltung, Projektplanung, Mailgruppe, Bugtracker

Möglichkeit

Distributed Mutation Database

Middleware

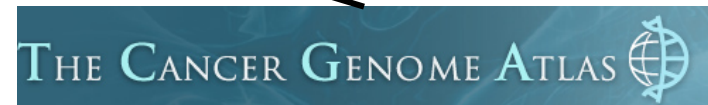
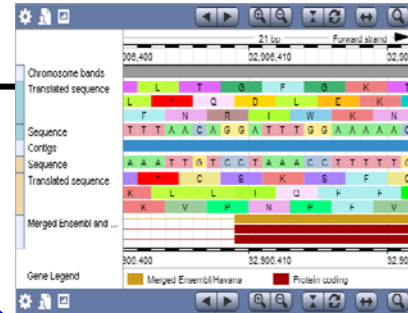


Möglichkeit

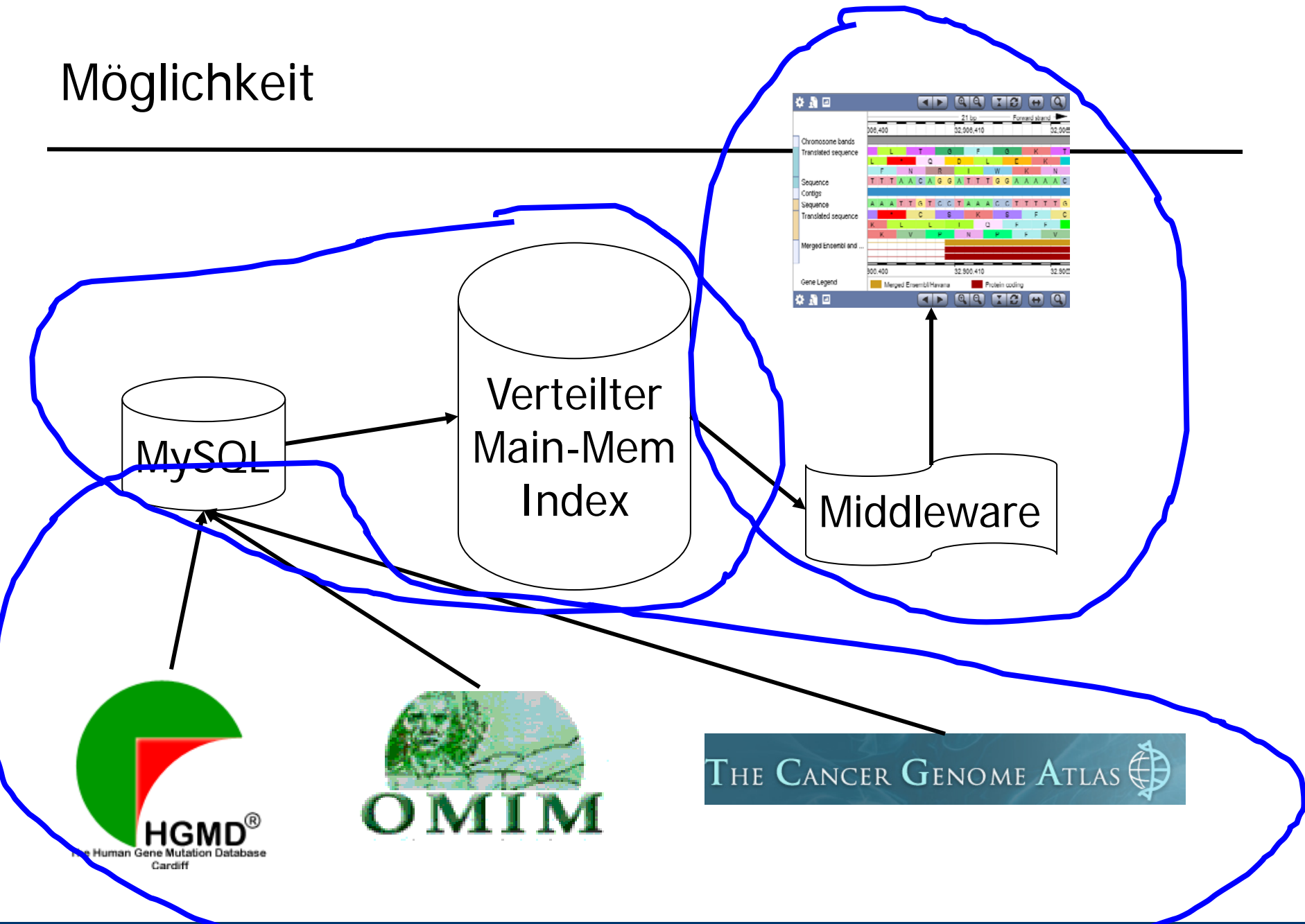
MySQL

Verteilter
Main-Mem
Index

Middleware



Möglichkeit



Use Case Beispiel

Ein Benutzer interessiert sich für Mutationen im Gen TP53 in Lungentumoren. Dazu sucht er in der Quelle TCGA nach allen Mutationen in diesem Bereich. Er sieht schnell, dass es sehr viele sind, und schränkt daher die Suche auf Mutationen mit einer gewissen Frequenz ein. Nun fragt er sich, ob diese Mutationen spezifisch für Lungentumore sind oder aber in jeder Bevölkerungsgruppe vorkommen. Daher wiederholt er die Suche, gibt nun aber TCGA und 1000 Genomes an. Beide Daten kann er in einer Visualisierung sehen und vergleichen, da beide Quellen in eigenen „Lanes“ über dem Referenzgenom angezeigt werden.



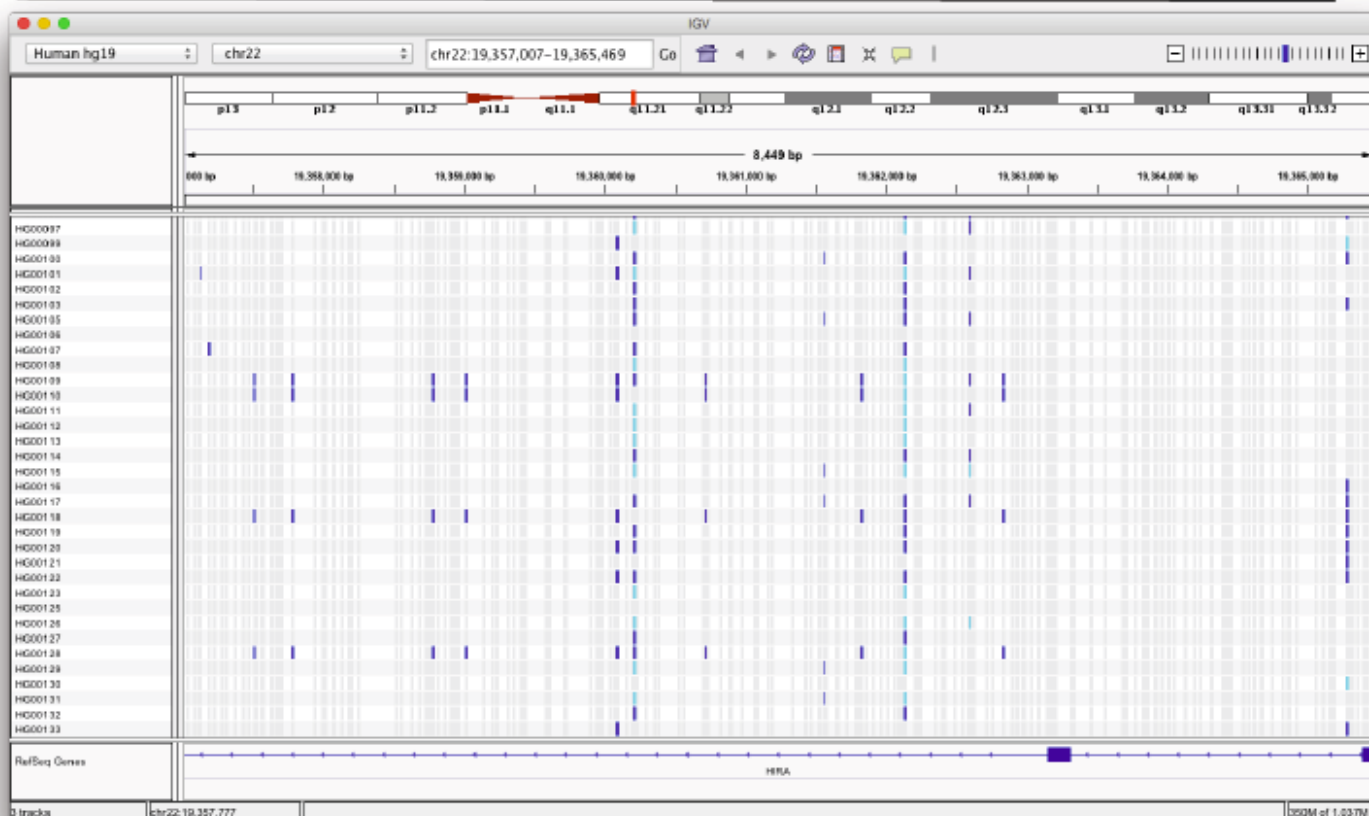
2 / 4



68,7%



Genombrowser: Integrative Genomics Viewer (IGV)





3

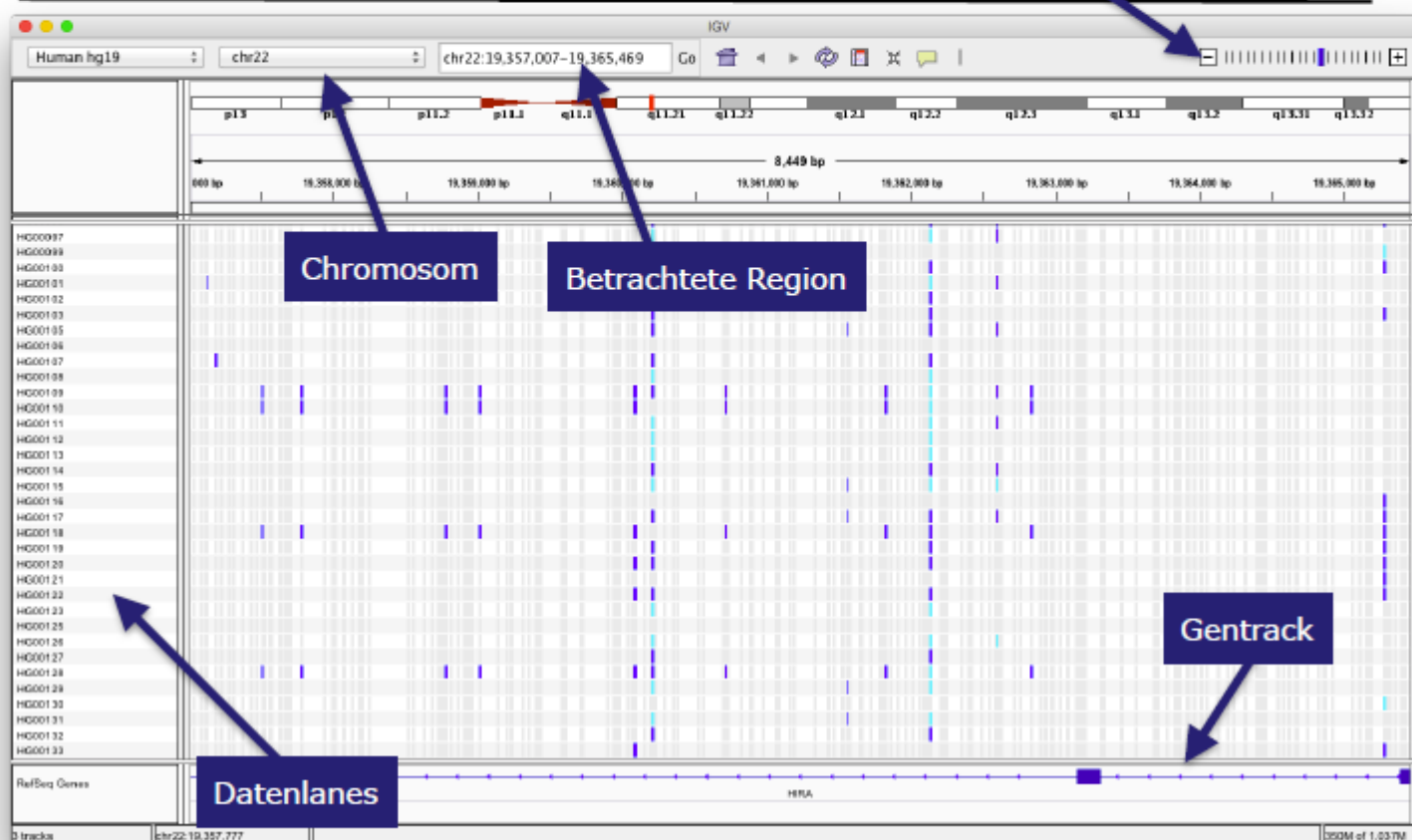
/ 4



68,7%



Genombrowser: IGV



Genombrowser: IGV

