



Jacob Asiedu



Josh Gould



Rajiv Narayan



Apps



Apps



Database



Apps



API



Database



Apps



Components



API



Database



Apps



Apps



Components



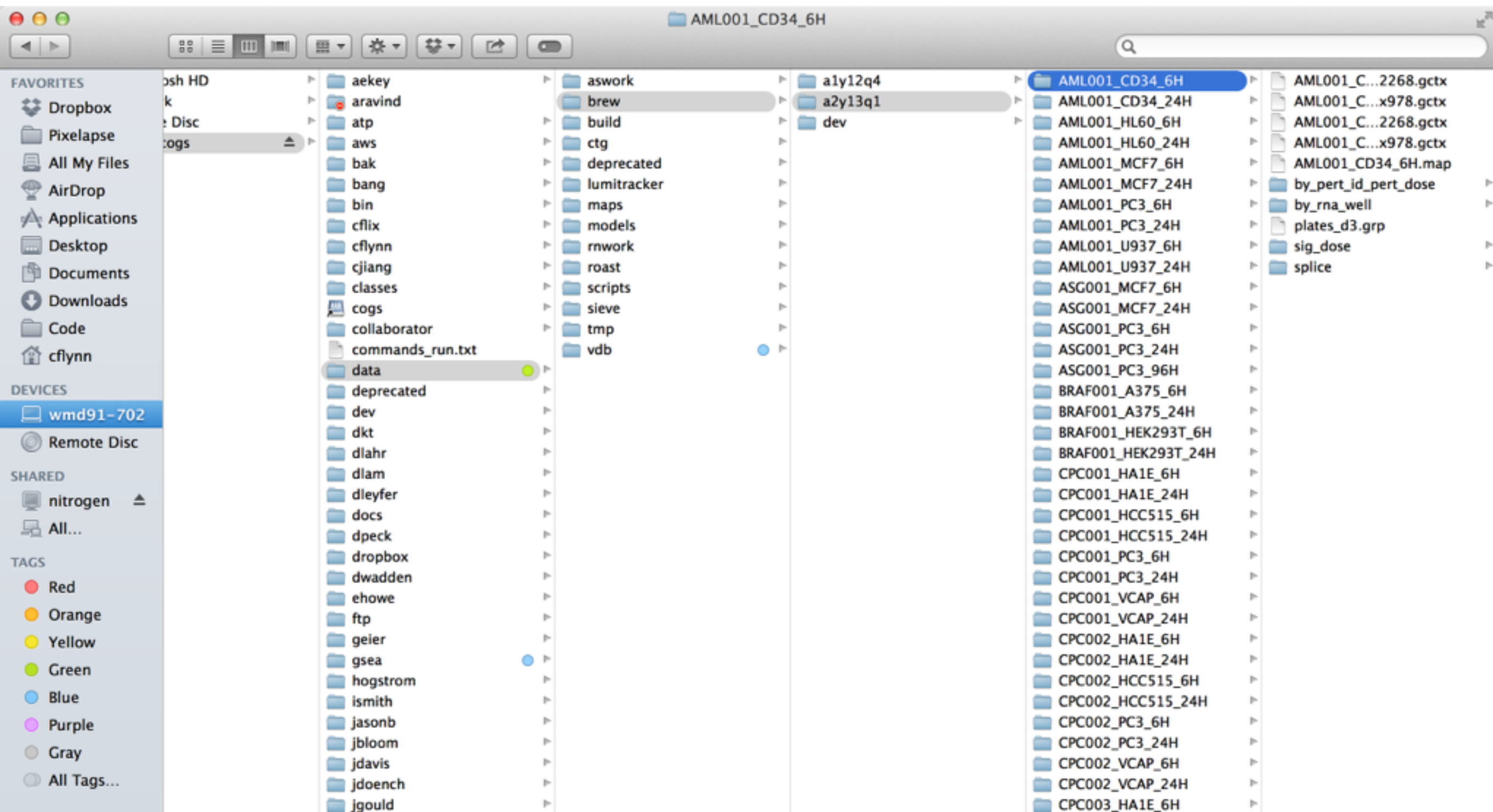
API



Database



File System





Database



mongoDB

Schemaless document store

Ideally suited for fast paced research environments



Database

```
2. mongo /Users/cflynn (mongo)
> db.evidence.find({"gene names":{"$regex":"myc","$options":"i"}})
{ "_id" : ObjectId("5489ee8d9e2d4a1eee194f14"), "ms/ms m/z" : 739.4, "acetyl (k) probabilities" : "LVSEK(1)LASYQAAR", "b mixture" : 17.395, "modified sequence" : "_LVSEK(ac)LASYQAAR_", "contaminant" : "", "m/z" : 739.4041, "pif" : 0, "calibrated retention time start" : 40.717, "intensity" : 754570000, "retention length" : 0.76836, "acetyl (protein n-term)" : 0, "peptide id" : 4882, "raw file" : "F20130816_TS_Titration_CSTKac_2mgJurkat_noXlink_25ug", "delta score" : 166.6, "id" : 17282, "uncalibrated mass error [ppm]" : -0.69282, "score" : 166.6, "combinatorics" : 1, "max intensity m/z 0" : 739.404653915951, "sequence" : "LVSEKLASYQAAR", "pep" : 1.2751e-11, "oxidation (m)" : 0, "ms/ms count" : 2, "charge" : 2, "experiment" : "Exp25", "fraction" : 1, "mass error [ppm]" : 0.11852, "retention time" : 40.926, "type" : "MULTI-MSMS", "aif ms/ms ids" : "", "oxidation (m) score diffs" : "", "fraction of total spectrum" : 0, "expID" : ObjectId("5489ee819e2d4a1eee190b91"), "acetyl (k) score diffs" : "LVSEK(166.6)LASYQAAR", "leading razor protein" : "P01106-2", "oxidation (m) site ids" : "", "calibrated retention time" : 40.926, "gene names" : "MYC", "calibrated retention time finish" : 41.485, "fasta headers" : [ ">sp|P01106-2|MYC_HUMAN Isoform 2 of Myc proto-oncogene protein OS=Homo sapiens GN=MYC", ">tr|H0YBT0|H0YBT0_HUMAN Myc proto-oncogene protein OS=Homo sapiens GN=MYC PE=4 SV=1", ">sp|P01106|MYC_HUMAN Myc proto-oncogene protein OS=Homo sapiens GN=MYC PE=1 SV=1", ">tr|B" ], "oxidation (m) probabilities" : "", "ms/ms scan number" : 11597, "protein names" : "Myc proto-oncogene protein", "acetyl (k) site ids" : 303, "leading proteins" : "P01106-2", "mod peptide id" : 5184, "modifications" : "Acetyl (K)", "reverse" : "", "base peak fraction" : 0, "proteins" : [ "P01106-2", "H0YBT0", "P01106", "B4E1N7", "H0YBG3" ], "uncalibrated - calibrated m/z [ppm]" : -0.81135, "retention time calibration" : 0, "ms/ms ids" : [ "18704", "18705" ], "length" : 13, "mass" : 1476.7936, "best ms/ms" : 18704, "match time difference" : "", "protein group ids" : 379, "resolution" : 40398.07, "acetyl (k)" : 1 }
{ "_id" : ObjectId("5489ee8d9e2d4a1eee194f15"), "ms/ms m/z" : 493.27, "acetyl (k) probabilities" : "LVSEK(1)LASYQAAR", "b mixture" : 17.395, "modified sequence" : "_LVSEK(ac)LASYQAAR_", "contaminant" : "", "m/z" : 493.27182, "pif" : 0, "calibrated retention time start" : 40.732, "intensity" : 280100000, "retention length" : 0.44383, "acetyl (protein n-term)" : 0, "peptide id" : 4882, "raw file" : "F20130816_TS_Titration_CSTKac_2mgJurkat_noXlink_25ug", "delta score" : 128.33, "id" : 17283, "uncalibrated mass error [ppm]" : -0.25236, "score" : 152.4, "combinatorics" : 1, "max intensity m/z 0" : 493.272422204878, "sequence" : "LVSEKLASYQAAR", "pep" : 8.8481e-10, "oxidation (m)" : 0, "ms/ms count" : 1, "charge" : 3, "experiment" : "Exp25", "fraction" : 1, "mass error [ppm]" : 0.18483, "retention time" : 40.926, "type" : "MULTI-MSMS", "aif ms/ms ids" : "", "oxidation (m) score diffs" : "", "fraction of total spectrum" : 0, "expID" : ObjectId("5489ee819e2d4a1eee190b91"), "acetyl (k) score diffs" : "LVSEK(152.4)LASYQAAR", "leading razor protein" : "P01106-2", "oxidation (m) site ids" : "", "calibrated retention time" : 40.926, "gene names" : "MYC", "calibrated retention time finish" : 41.175, "fasta headers" : [ ">sp|P01106-2|MYC_HUMAN Isoform 2 of Myc proto-oncogene protein OS=Homo sapiens GN=MYC", ">tr|H0YBT0|H0YBT0_HUMAN Myc proto-oncogene protein OS=Homo sapiens GN=MYC PE=4 SV=1", ">sp|P01106|MYC_HUMAN Myc proto-oncogene protein OS=Homo sapiens GN=MYC PE=1 SV=1", ">tr|B" ], "oxidation (m) probabilities" : "", "ms/ms scan number" : 11599, "protein names" : "Myc proto-oncogene protein", "acetyl (k) site ids" : 303, "leading proteins" : "P01106-2", "mod peptide id" : 5184, "modifications" : "Acetyl (K)", "reverse" : "", "base peak fraction" : 0, "proteins" : [ "P01106-2", "H0YBT0", "P01106", "B4E1N7", "H0YBG3" ], "uncalibrated - calibrated m/z [ppm]" : -0.81135, "retention time calibration" : 0, "ms/ms ids" : [ "18704", "18705" ], "length" : 13, "mass" : 1476.7936, "best ms/ms" : 18704, "match time difference" : "", "protein group ids" : 379, "resolution" : 40398.07, "acetyl (k)" : 1 }
```



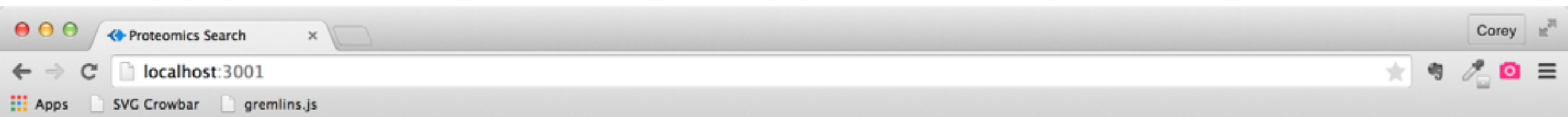

API

```
localhost:3000/search?q=
localhost:3000/search?q={"gene%20names":{"$regex":"myc","$options":"i"}}
Apps SVG Crowbar gremlins.js
View source

{
  - evidence: [
    - {
      _id: "5489ee8d9e2d4aleee194f14",
      ms/ms m/z: 739.4,
      acetyl (k) probabilities: "LVSEK(1)LASYQAAR",
      b mixture: 17.395,
      modified sequence: "_LVSEK(ac)LASYQAAR_",
      contaminant: "",
      m/z: 739.4041,
      pif: 0,
      calibrated retention time start: 40.717,
      intensity: 754570000,
      retention length: 0.76836,
      acetyl (protein n-term): 0,
      peptide id: 4882,
      raw file: "F20130816_TS_Titration_CSTKac_2mgJurkat_noXlink_25ug",
      delta score: 166.6,
      id: 17282,
      uncalibrated mass error [ppm]: -0.69282,
      score: 166.6,
      combinatorics: 1,
      max intensity m/z 0: 739.404653915951,
      sequence: "LVSEKLASYQAAR",
      pep: 1.2751e-11,
      oxidation (m): 0,
      ms/ms count: 2,
      charge: 2,
      experiment: "Exp25",
      fraction: 1,
      mass error [ppm]: 0.11852,
      retention time: 40.926,
      type: "MULTI-MSMS",
      aif ms/ms ids: "",
      oxidation (m) score diffs: "",
      fraction of total spectrum: 0,
      expID: "5489ee819e2d4aleee190b91",
      acetyl (k) score diffs: "LVSEK(166.6)LASYQAAR",
      leading razor protein: "P01106-2",
      oxidation (m) site ids: "",
```



Components

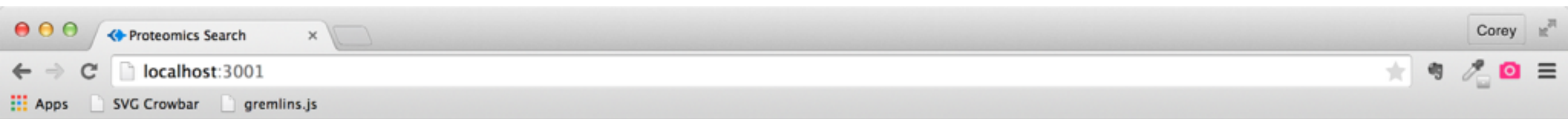


Sequence	Modifications	Count	Total Intensity	Average Intensity
LVSEK(ac)LASYQAAR	Acetyl (K)	10	5086410000	508641000
DYPAAK(ac)R	Acetyl (K)	3	55484000	18494666.666666668
HHLGAATPENPEIELLR	Unmodified	5	13656000	2731200

tsv csv png



Apps



MYC

File Path

/home/ec2-user/TitrationCurve

Sequence	Modifications	Count	Total Intensity	Average Intensity
LVSEK(ac)LASYQAAR	Acetyl (K)	10	5086410000	508641000
HHLGAATPENPEIELLR	Unmodified	5	13656000	2731200
DYPAAK(ac)R	Acetyl (K)	3	55484000	18494666.666666668

tsv csv png





Database

//lime/silac/Jake/exp1/evidene.txt

path

collection

M
O
N
G
O

experiments →

peptides

modSpecPeptides

proteinGroups

evidence →

[{ id: 123, path: "...Jake/exp1/" }
{ id: 456, path: "...Jake/exp2/" }]

[{ expID: 456, genes: [A] },
{ expID: 456, genes: [B] }]



Database

total file size on lime: 8.0 TB

total file size in mongo: 526 GB

number of experiments: 1326

number of files ingested: 6630

number of collections: 6

number of documents: 186,242,497