

## Supplementary Data and Analysis

Hill, J. I.; MacMullen, W. J.; & Palmer, C. L. Characteristics of bioinformatics employment advertisements. *Proceedings of the 2009 Annual Meeting of the American Society for Information Science & Technology*.

This file provides additional data and analysis for which space was not available in the published paper. The authors thank Mijung Yoon for her assistance.

Tables S1-S3. Skills requested for master's level by domain in each of 3 main employment settings (selected 2007-2008 data (n=404))

Table S1. CS skills

CS Skills	Major Employment Setting			Total
	Academic	Corporate	Research Institute	
Programming	6	26	15	47
Oracle+Databases	6	22	15	43
Unix+Linux	6	15	8	29
Web	5	4	10	19
Software development	1	11	5	17
Data mining	1	1	0	2
Algorithms	1	6	1	8
XML	1	0	3	4
Version control systems	0	1	0	1
Web service architecture	0	0	2	2
Mac OS	0	1	0	1
No tech skills specified	6	7	1	14

Table S2. Bio/info skills

Bio/info skills	Major Employment Setting			Total
	Academic	Corporate	Research Institute	
BI tools	4	13	7	24
BI resources	4	9	4	17
Sequence analysis	2	11	5	18
Microarray	1	3	3	7
Genetics/genomics	2	9	1	12
Molecular bio	1	6	0	7
Lab experience	1	0	0	1
Pharma	0	3	0	3
Proteomics	0	2	0	2
Clinical research	0	3	0	3
Biomed	0	5	1	6
Phylogenetic tools	1	1	0	2
Genome annotation	0	3	0	3
Statistics	2	16	1	19

Table S3. All skills

All Skills	Major Employment Setting			Total	Average
	Academic	Corporate	Research Institute		
Programming	6	26	15	47	15.7
Oracle+Databases	6	22	15	43	14.3
Unix+Linux	6	15	8	29	9.7
BI tools	4	13	7	24	8.0
Web	5	4	10	19	6.3
Statistics	2	16	1	19	6.3
Sequence analysis	2	11	5	18	6.0
Software development	1	11	5	17	5.7
BI resources	4	9	4	17	5.7
No tech skills specified	6	7	1	14	4.7
Genetics/genomics	2	9	1	12	4.0
Algorithms	1	6	1	8	2.7
Microarray	1	3	3	7	2.3
Molecular bio	1	6	0	7	2.3
Biomed	0	5	1	6	2.0
XML	1	0	3	4	1.3
Pharma	0	3	0	3	1.0
Clinical research	0	3	0	3	1.0
Genome annotation	0	3	0	3	1.0
Data mining	1	1	0	2	0.7
Web service architecture	0	0	2	2	0.7
Proteomics	0	2	0	2	0.7
Phylogenetic tools	1	1	0	2	0.7
Version control systems	0	1	0	1	0.3
Mac OS	0	1	0	1	0.3
Lab experience	1	0	0	1	0.3
LIMS	0	0	0	0	0.0

\* Skills are listed by the number of listings for all three settings from high to low.

Figures S1-3. Skill set profiles for master's level in each of 3 main employment settings (selected 2007-2008 data (n=404))

Figure S1. Academic setting

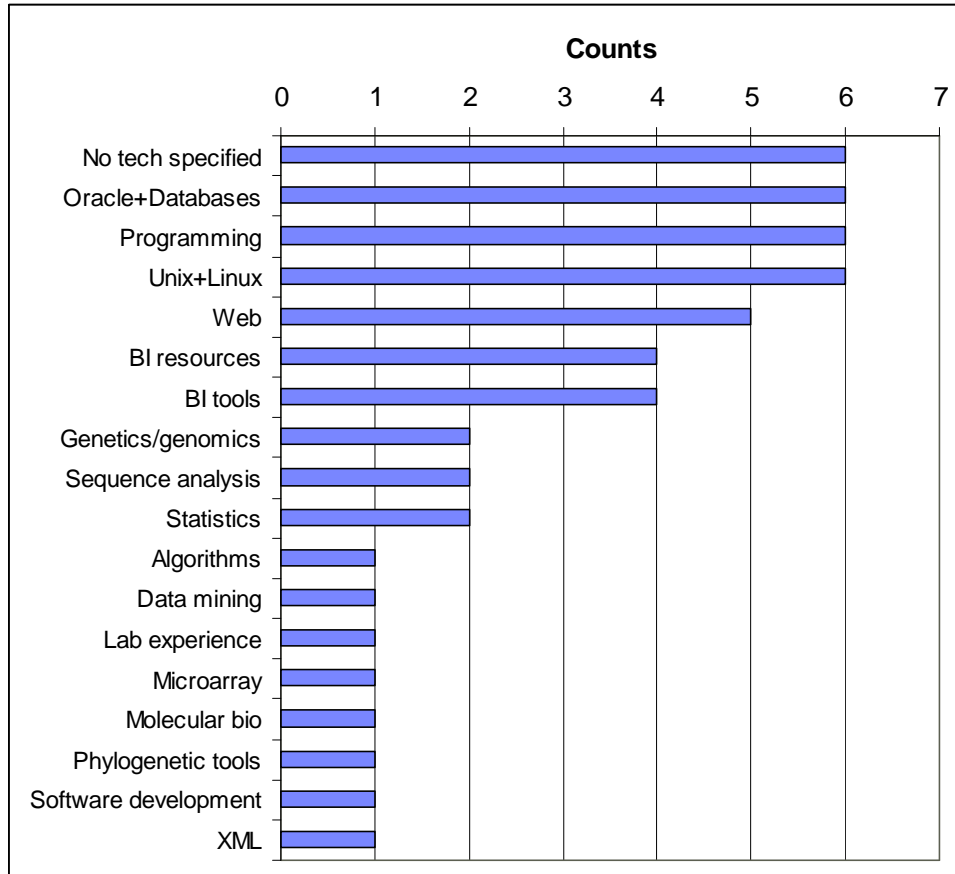


Figure S2. Corporate setting

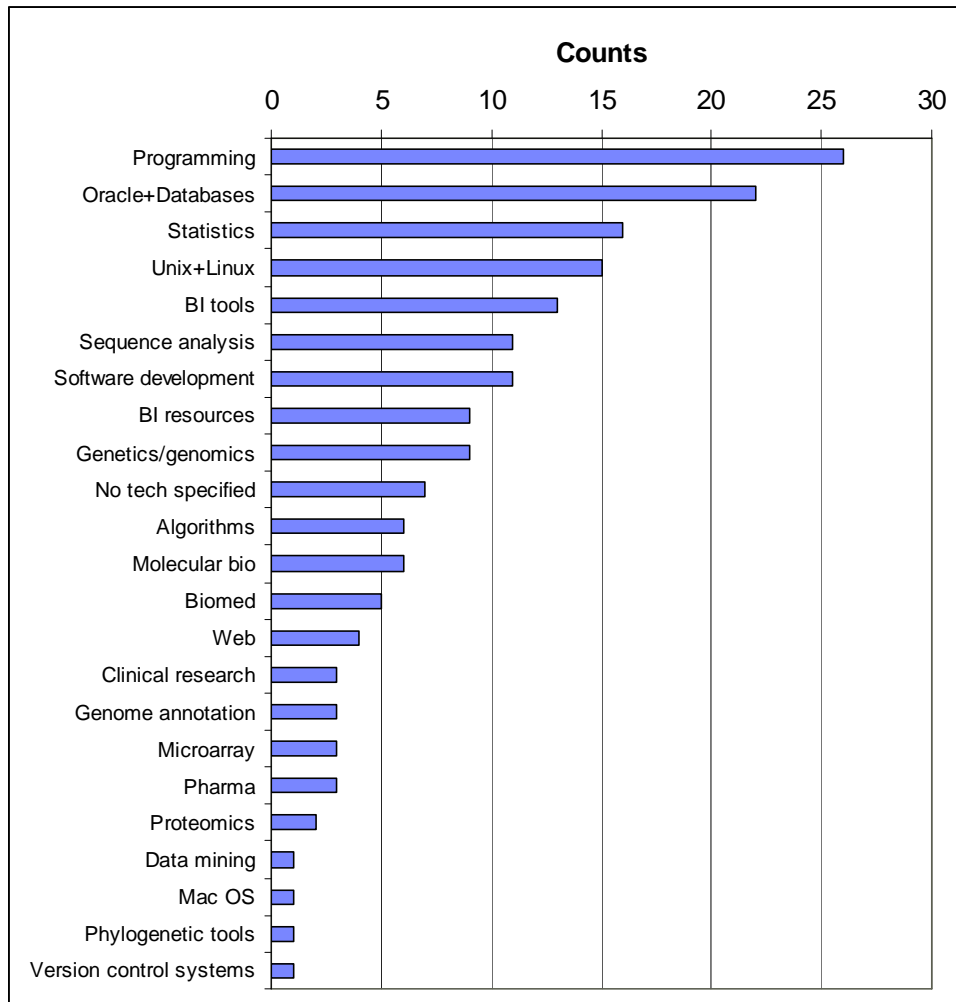
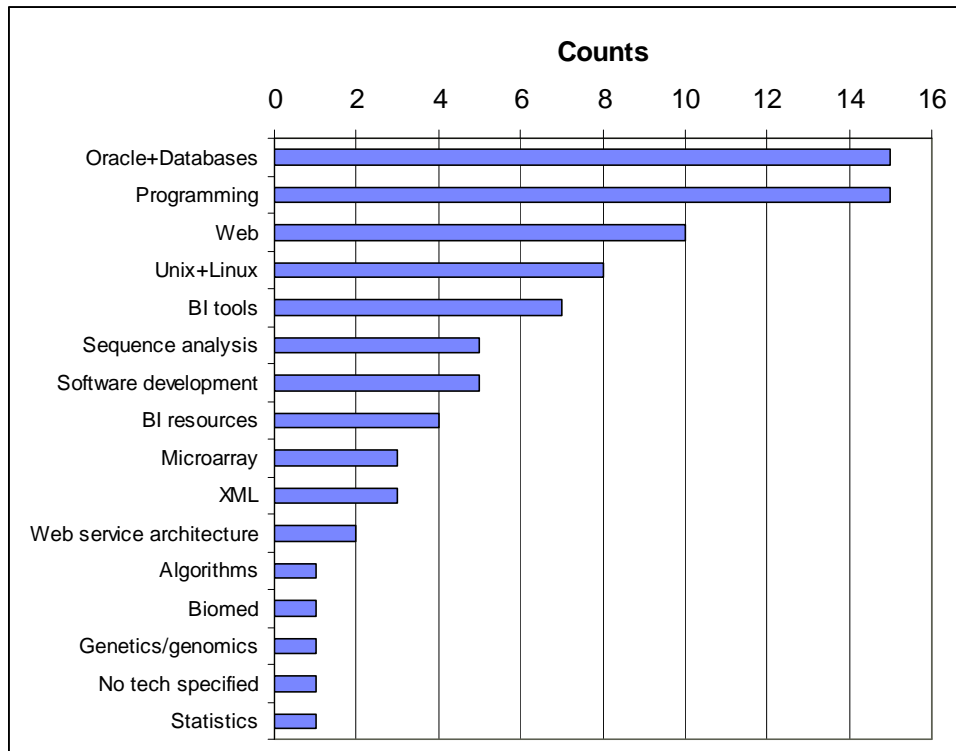


Figure S3. Research Institute setting



## Summary of Position Descriptions: Programmer

There were 66 positions in the 2007-2008 data set (n=404). The most common responsibility associated with a programmer position was to develop and support bioinformatics software and databases for the institution-wide use or for specific research projects, often working closely with researchers and scientists. The programmer positions were often titled as Bioinformatics Programmer, Scientific Programmer, Programmer Analyst, or Programmer. While most of the programmer positions shared the common responsibility of computational programming support for the institution or projects, programmer analyst positions sometimes included additional responsibilities of data transfer and modeling. The following is a summary of the 66 programmer positions:

1. Development, modification, and maintenance of bioinformatics databases and software.
  - Developing tools for browsing, managing and integrating research/clinical data or databases, including:
    - Designing and writing databases and software for biological and clinical data management (e.g., defining architecture of customized databases for the institution's infrastructure);
    - Testing the software by performing edit checks in compliance with standard operating procedures and regulatory agency guidelines, in coordination with data management;
    - Documenting and reporting data management process and outcome, and producing dictionary-coding programs and miscellaneous data management listings;
    - Creating database conversion programs;
  - Providing support for implementation and maintenance of custom or third-party bioinformatics software or databases, including:
    - Operating and upgrading software and databases for data management or analysis (e.g., web- or report-based sequence mapping tools);
    - Troubleshooting software problems;
  - (often titled as a scientific/research programmer) providing programming support for scientists in architecture, design, implementation, and testing of software application suites used for collecting, analyzing, reporting, and sharing (genetic and genomic) data, including
    - Maintaining close communication regarding data management programming activities with researcher(s);
    - Implementing tools for data modeling, analysis, and interpretation;
    - Developing and applying statistical techniques to data analysis (e.g., identification of genes for diseases);
    - Implementing middleware;
2. (often with the term 'systems' in a position title) building and maintaining infrastructure for computational biology, including: (1) development of a hardware platform; (2) configuration, maintenance, and integration of servers, workstations, clusters, and advanced storage under multiple OSes; (3) interfacing with network configuration and

- security; (4) developing reliable procedures for data backup and restoration; (5) helping define and implement user policies; and (5) implementing new system tools/processes.
3. (often titled as programmer analyst or programmer/data analyst) performing data analysis (e.g., microarray data analysis, sequence mapping);
  4. Developing web resources, improving websites and managing information portals
  5. (research programmer) data modeling, database design, middleware implementation, implementation of web and data services.

Terms That Were Frequently Associated with 'Programmer'  
in Job Titles in the 2007-2008 data (n=404)

Terms	Frequency
Bioinformatics, computer, or no additional terms	23
Analyst	16
Scientist/scientific/research	12
Systems /systems administrator/ systems analyst	4
Database administrator/application	4
Specific areas (e.g., immunology, plant genomics, clinical)	4
specific software (e.g., perl, oracle, java)	4
Computational Biology	3
Data analyst	2
Biostatistics	1
Web	1
TOTAL	74

\* The total doesn't equal to the number of positions due to several listings that had multiple terms that were associated with 'programmer'.

## Summary of Position Descriptions: Curator

This description is based on the ten curator positions that were listed between January 2007 and June 2008. The descriptive terms listed in the titles along with the term 'curator' included: scientific, genome, microarray, genotype, phenotype, and name of a database (e.g., dbGaP, dbSNP). Qualifications sought for the position included: education (e.g., molecular biologist), background/expertise (e.g., life science, genotype analysis). Most of the positions listed responsibilities related to curation, review, and processing records for analysis. Less common responsibilities included literature review, data analysis, and communication with investigators. The detailed summary of responsibilities is as follows:

1. Ensuring data quality
  - a. Define data (e.g., genomic sequence records) standards
2. Data curation and processing (e.g., perform "advanced curation" of data into comparable datasets)
  - a. Reviewing and processing incoming data/records submissions
  - b. Developing infrastructure for data analysis (e.g., applications to process data into databases and CGI)
  - c. Using specialized databases and bioinformatics software
  - d. Providing functional annotation with records and companion resources
3. Perform data analysis
  - a. Computational analysis using various bioinformatics and programming tools
  - b. Scholarly interpretation of results
4. Documentation
  - a. Annotation and integration of study meta data (e.g., protocols and data collection forms)

Examples of qualifications listed for curator positions:

Computational biologist with background in genotype analysis

Biologist with expertise in life sciences

Genomic scientist with background in sequence analysis and positional cloning

Molecular biologist/biochemist