1. What is the RNA source (and corresponding experiment ID) for all of the expression experiments?

2. SELECT * FROM Descriptions LIMIT 10;

3. How many (different) LocusLinks are in the database?
   
   How does this compare to SELECT count(linkId) FROM LocusLinks; ?  
   Ans: Both should give the same result, since linked is the primary key.

4. Either of these is correct:
   
   SELECT count(*) from Targets;  
   SELECT count(affyId) from Targets;

5. What’s the description and species of LocusLink 655?

6. SELECT level from Data  
   WHERE affyID = "33659_at"  
   AND exptId = "hs-liv-1";

7. What are all the LocusLink descriptions, along with their IDs, starting with the word “phosphatase”?

8. SELECT * FROM Descriptions  
   WHERE description LIKE "%growth factor%";

9. What data points in the experiment “mm-hrt-1” have expression > 6000? Make the list in order of decreasing expression.

10. SELECT * FROM GO_Descr  
    WHERE description LIKE "%transcription factor%"  
    ORDER BY description ASC;

11. What ten data points show the highest expression levels? In your output include Affy ID, corresponding GenBank ID of the target sequence, and expression level.
12. SELECT Targets.affyId, 
    Targets.gbId, 
    Descriptions.description 
FROM Targets, Descriptions 
WHERE Descriptions.description LIKE "%interleukin%" 
    AND Descriptions.gbId = Targets.gbId;

13. What ten Affy IDs with a target sequence belonging to a Unigene cluster show the highest expression level? Report your answer including Affy ID, Target GenBank ID, Unigene ID, and expression level. Note that any Affy IDs that don’t belong to a Unigene cluster (i.e., without a uId) will not appear in the list. This is an example where it’s important to pay attention to the cardinalities in your E-R diagram.

14. SELECT Targets.affyId, UniSeqs.uid, UniDescr.description 
FROM Targets, UniSeqs, UniDescr 
WHERE Targets.gbId = UniSeqs.gbId 
    AND UniSeqs.uId = UniDescr.uid 
ORDER BY UniDescr.description DESC 
LIMIT 10;

15. What is the average expression level of ten genes (Affy IDs) showing the highest average expression across all six experiments?

16. SELECT exptId, AVG(level) AS average 
FROM Data 
GROUP BY exptId;

17. What Affy IDs show expression in one experiment that’s at least 10 times the expression in another experiment. Skip the control targets from the analysis.

18. SELECT Data.affyId 
FROM Data, Data DataCopy 
WHERE Data.level > DataCopy.level 
    AND Data.affyId=DataCopy.affyId 
    AND Data.exptId = "hs-liv-1" 
    AND DataCopy.exptId = "hs-hrt-1" 
    AND Data.affyId NOT LIKE "AFFX%" 
LIMIT 10;

19. What ten affy IDs (along with level in heart, level in brain, and the difference) show the greatest difference in expression levels in these tissues?

20. This is a surprisingly difficult query in MySQL, since nested queries are not yet allowed. You may not have permission to create tables this week.
# Create a temporary table to hold Affy IDs and maximum values
# Use a name like your own name so others won’t try to access the same table.
CREATE TEMPORARY TABLE myname
(
  affyId varchar(30) NOT NULL,
  level int(11) unsigned
);

# Select Affy IDs and maximum values and put them into this new table
INSERT INTO myname
SELECT affyId, MAX(level)
FROM Data GROUP BY affyId;

# Find rows in Data that match level of rows in temporary table (and print the top 100)
SELECT Data.affyId, Data.exptId, Data.level
FROM Data, myname
WHERE Data.affyId = myname.affyId
  AND Data.level = myname.level
ORDER BY Data.level DESC LIMIT 100;

# Delete the temporary table
DROP TABLE myname;

21. Same as question (19), but also list LocusLink ID and description for each Affy ID.