What is claimed is:

1. A computer-implemented method of identifying rare variants that cause extreme levels of gene expression, including:

accessing gene expression levels for a group of individuals;

normalizing the gene expression levels, and identifying those outlier individuals from the group of individuals that have extreme levels of gene expression, wherein the extreme levels of gene expression are determined from tail quantiles of the normalized gene expression levels;

selecting rare variants from gene sequences of the outlier individuals, wherein the rare variants are selected based on an allele frequency cutoff;

fitting a causality model to determine causal relationships between the rare variants and the extreme levels of gene expression in the outlier individuals while controlling for a plurality of confounders; and

generating causality scores for the rare variants based on the determined causal relationships, wherein a particular causality score of a particular rare variant indicates a likelihood of the particular rare variant causing an extreme level of gene expression in those outlier individuals whose gene sequences contain the particular rare variant.

2. The computer-implemented method of claim 1, wherein the causality scores are probability values (p-values).

3. The computer-implemented method of claim 1, wherein the p-values are determined by a Pearson correlation coefficient.

4. The computer-implemented method of claim 1, wherein the causality model is a logistic regression model, a linear regression model, an analysis of covariance (ANCOVA) model, and/or a multivariate analysis of covariance (MANCOVA) model.

5. The computer-implemented method of claim 4, wherein the fitted causality model determines the causal relationships by predicting a particular gene expression level of a particular gene in a particular chromosome in dependence upon a variant-driven gene expression level caused by a particular rare variant.

6. The computer-implemented method of claim 1, wherein the plurality of confounders includes distal *trans*-expression quantitative trait loci (eQTLs) effects.

7. The computer-implemented method of claim 6, wherein the fitted causality model controls for the distal *trans*-eQTLs effects by predicting the particular gene expression level in dependence upon a *trans* gene expression level caused by other genes in other chromosomes.

8. The computer-implemented method of claim 1, wherein the plurality of confounders includes local *cis*-eQTLs effects.

9. The computer-implemented method of claim 1, wherein the plurality of confounders includes population structure and ancestry effects.

10. The computer-implemented method of claim 1, wherein the plurality of confounders includes probabilistic estimation of expression residuals (PEER) effects.

11. The computer-implemented method of claim 1, wherein the plurality of confounders includes environmental effects.

12. The computer-implemented method of claim 1, wherein the plurality of confounders includes gender effects, batch effects, genotyping platform effects, and library construction protocol effects.

13. The computer-implemented method of claim 1, wherein the extreme levels of gene expression include over gene expression and under gene expression.

14. The computer-implemented method of claim 1, wherein the rare variants are non-coding variants.

15. The computer-implemented method of claim 1, wherein the gene expression levels are further stratified into tissue-specific gene expression levels for a plurality of tissues.

16. The computer-implemented method of claim 1, wherein the causality model is fitted using stratification.

17. The computer-implemented method of claim 1, further including generating a ranking of the rare variants based on the causality scores.

18. The computer-implemented method of claim 1, wherein the rare variants are singleton variants.

19. A system including one or more processors coupled to memory, the memory loaded with computer instructions to identify rare variants that cause extreme levels of gene expression, the instructions, when executed on the processors, implement actions comprising:

accessing gene expression levels for a group of individuals;

normalizing the gene expression levels, and identifying those outlier individuals from the group of individuals that have extreme levels of gene expression, wherein the extreme levels of gene expression are determined from tail quantiles of the normalized gene expression levels;

selecting rare variants from gene sequences of the outlier individuals, wherein the rare variants are selected based on an allele frequency cutoff;

fitting a causality model to determine causal relationships between the rare variants and the extreme levels of gene expression in the outlier individuals while controlling for a plurality of confounders; and

generating causality scores for the rare variants based on the determined causal relationships, wherein a particular causality score of a particular rare variant indicates a likelihood of the particular rare variant causing an extreme level of gene expression in those outlier individuals whose gene sequences contain the particular rare variant.

20. A non-transitory computer readable storage medium impressed with computer program instructions to identify rare variants that cause extreme levels of gene expression, the instructions, when executed on a processor, implement a method comprising:

accessing gene expression levels for a group of individuals;

normalizing the gene expression levels, and identifying those outlier individuals from the group of individuals that have extreme levels of gene expression, wherein the extreme levels of gene expression are determined from tail quantiles of the normalized gene expression levels;

selecting rare variants from gene sequences of the outlier individuals, wherein the rare variants are selected based on an allele frequency cutoff;

fitting a causality model to determine causal relationships between the rare variants and the extreme levels of gene expression in the outlier individuals while controlling for a plurality of confounders; and

generating causality scores for the rare variants based on the determined causal relationships, wherein a particular causality score of a particular rare variant indicates a likelihood of the particular rare variant causing an extreme level of gene expression in those outlier individuals whose gene sequences contain the particular rare variant.