Structural and Operational Complexity of the *Geobacter sulfurreducens* Genome

Yu Qiu[a], Byung-Kwan Cho[a], Young Seoub Park[a], Derek Lovley[b], Bernhard Ø. Palsson[a], and Karsten Zengler[a,1]

[a]Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA.

[b]Department of Microbiology, University of Massachusetts, Amherst, MA 01003, USA.

SI Figure Legends

Figure S1. Examples of open reading frame (ORF) correction based on peptide reads, expression profile, transcription start site (TSS), and start codon prediction. (*A, B*) Examples show corrected ORFs (orange arrows) from current annotation (grey arrows) based on expression profiles and TSSs that were determined downstream of current start codon (marked by asterisk). ORFs were shortened to a new start codon in frame (shown in red).

Figure S2. Confirmation of six randomly selected sRNAs by northern blot. Length of experimentally annotated sRNAs was confirmed by northern blot (marker length in base pairs).

Figure S3. Comparison of 5'UTR length (in base pairs). No difference in 5'UTR length between different functional categories was observed.

Figure S4. Genome coverage by iterative integration of transcriptome data. Data from the following five growth conditions were used: (1) growth in fumarate-acetate medium, mid-logphase; (2) growth in fumarate-ferric citrate medium, mid-logphase; (3) growth in modified fumarate-acetate medium, with N_2 as nitrogen source, mid-logphase; (4) growth with fumarate-acetate medium as a biofilm; and (5) growth on electrode in a microbial fuel cell.