Examiner 1 comments

NOTE: Examiner's comments in black and candidate's comments in red.

General comments:

The study examines three facial dysostoses - Treacher Collins, Nager and Miller syndromes — with the aim of using a targeted next-generation sequencing approach to generate a mutation profile for South African patients. The candidate successfully generated targeted NGS data as part of a larger study on developmental disorders. Fifteen patients were recruited and seven putative disease-causing variants identified in seven of the 15 patients, a diagnostic yield of 47%. The results indicate that NGS may be a useful approach in confirming a suspected clinical diagnosis for these overlapping conditions. This is the first study of its kind on South African patients.

Overall, the dissertation is well presented and logically laid out. There are typos and grammatical errors which could have been avoiding by using spelling and grammar checks. Try to avoid repetition within a sentence. A few cases of word-for- word repetition I believe in an attempt to emphasize a point, can also be avoided. Tables and figures are appropriately used. The final version should be carefully checked and all errors corrected. The dissertation has been checked and corrected for typos, grammatical errors and repeated statements have been removed throughout the document.

The introduction provides a good review of key references and background to the study. Although the three FDs mentioned above are presented, patient recruitment and data analysis includes additional conditions with overlapping features. The differential diagnoses are only briefly mentioned in Section 1.7 under study rationale. Two deceased patients were included in this study but the consent and DNA samples used are not described anywhere; this is important information that needs to be included. Check supplier vs manufacturer for reagents and equipment used. Check consistency of abbreviations (must be written in full the first time) and numbers (some are written in words and others not). To address the concern of differential diagnoses disorder only briefly mentioned in Section 1.7,

A paragraph on the differential diagnoses of the three disorders is included in the introduction section; see page 15, para 1-2

A paragraph addressing the concern regarding consent of deceased patients has been included under the Materials and Methods section, Page 22, Para 1

All abbreviations have been written in full when mentioned for the first time in the dissertation. Numbers less than and equal to ten are written in words and greater than ten in numerals throughout the dissertation throughout the dissertation.

All reagents and equipment used were checked for supplier vs manufacturer.

The study results are presented in appropriate detail. Figure and table legends are missing some abbreviations and vice versa (abbreviations in the legend are not in the table). Check that all referrals to tables and/or figures within accompanying text are correct. Check ALL figure legends for consistency when referring to (a) and (b). Gene names must be in italics - particularly lacking in figure legends. For patient photographs, (a) and (b) should be placed outside of the image for better visibility. This needs to be corrected throughout.
Abbreviations in all Figure and Table legends have been checked and corrected.
All figure legends were checked and corrected for consistency when referring to (a) and (b) labels.
The gene names have been italicised in all Figure legends.
Regarding concern for visibility of patient photos, the labels (a) and (b) have been placed in a white space outside the images.

The discussion provides a clear and detailed analysis of the putative disease-causing mutations identified in this study within the context of disease physiology and the likely mechanism of action. This section in particular would benefit from proof- reading and editing to correct typos, grammar and phrasing before final submission, The CHARGE syndrome phenotype is covered in the discussion although only briefly mentioned in the Introduction as a differential diagnosis — rather address the related literature in the

introduction rather since 5/15 patients recruited were clinically diagnosed with CHARGE syndrome (with one additional query) AND four were confirmed. Although the sample size is small, this is a good pilot project investigating the application of targeted NGS for a group of genetic conditions, in a diagnostic setting with limited resources. The results are a significant first step towards developing a mutation profile for South African patients affected by facial dysostoses. The candidate has shown her ability to perform and present the laboratory work, data analysis and interpretation of variants. Study limitations were addressed and some recommendations have been made.

The section has been proof-read and edited for grammatical and spelling errors.

With regard to the concern that CHARGE syndrome was only briefly mentioned in the Introduction section, a paragraph on the three differential diagnoses disorder has been added to the Introduction section, see page 15, para 1-2.

Preliminary pages

1. The list of appendices begins with Appendix 'one', the next is Appendix B1. Please correct for consistency. The list of appendices has been corrected.
2. Xx and xi. The majority of gene names within the list of figures are not in italics. Check the rest of the dissertation and correct. All gene names in the dissertation are italicized.
3. Adjust the spacing in the list for Figure 3.8; the numbers make it difficult to distinguish. Space is adjusted.
4. xiii, Abbreviations: INDELS is listed twice in the table (INDELS and Indels on the next page). ExAc is listed but does not appear within the dissertation. Edit tRNA (extra 'tRNA' at the end) and correct 'VEST', should be VEP. The repeated 'indel' word is removed from the list. Transfer RNA (tRNA) has only been listed once and the candidate used the VEST tool in her analysis not VEP.

Introduction

5. Page 1: The reference for Fig 1.1 (g) - Wieczoreck, 2013 — is misspelt throughout the dissertation. The correct spelling Wieczorek (without the 'c' at the end) is in the reference list. Fig 1.1 has been removed from the dissertation and the correct "Wieczorek" spelling has been used throughout the dissertation.
6. As mentioned in general comments, the letters within the figures are unclear in some cases and should rather be in the white space. The labelling of figures has been corrected as suggested.
7. Page 2: The source of the image must be stated; assume it is taken from the reference (Hamilton and Mossman, 1972). Hamilton and Mossman is not in the reference list — it should be included. The image has been removed from the dissertation as permission to reproduce was not obtained.
8. Page 3: The source of the image is not stated; assume it is taken from the reference in the legend (Lizier et al. 2013) — also not in the reference list and | am unable to check. The image has been removed from the dissertation as permission to reproduce was not obtained.
9. Page 4: A whole paragraph is used to describe MFDs and AFDs but there is no final statement about whether or not TCS, MS and NS fit into either category or not. A sentence describing the category of TCS, MS and NS has been added and reads as "While there are at least eight different MFD's and eighteen different AFDs reported in literature, the present study focused on three most common and well understood FDs, namely, one MFD (Treacher Collins syndrome (TCS) and two AFDs (Miller syndrome (MS) and Nager syndromes (NS))". Page 2, para 1
10. Do not use apostrophes (') for plural terms i.e. AFDs NOT AFD's. Also correct RNA's on page 10 and check use of abbreviations elsewhere in the dissertation. Corrected throughout the dissertation.
11. Do not start a sentence with an abbreviation, the second paragraph begins with 'TCS' — must be written in full at the start of a sentence. This applies to many other areas in the dissertation (e.g. 'TCS' on page 6 multiple times; 'MS' on page 12, etc.), please review all and correct. Reviewed and corrected throughout the dissertation.

12. Paragraph 2: “However, these are global statistics and may not reflect the South African situation with respect to TCS, MS and NS prevalence rates, as no molecular and phenotypic studies...” Highlighted portion of this sentence is unnecessary repetition as the paragraph presents statistics. The statement has been removed. Page 2, para 2
13. Paragraph 3: “Despite this, a molecular confirmation of diagnosis can still have a positive effect on a patient's life as research shows that individuals who undergo genomic sequencing for rare disorders report interest in receiving results for reasons beyond clinical utility.” Please re-write this sentence, is currently part plural and part singular. The statement has been rephrased. Page 2, Paragraph 3
14. Page 5: I am not familiar with the use of “reviewed in...” within the reference bracket. Please reconsider and correct if appropriate. “reviewed in” was used to reference a review article, however, its use has been reconsidered and removed from all the references throughout the dissertation.
15. Page 6: “TCS was first described by Thomson, Toynbee and Berry independently in the 1880s (Thomson, 1846; Toynbee, 1847; Berry, 1889).”Two of the authors published in the 1840s. The statement has been corrected. Page 4, Paragraph 1
16. Fig 1.4: image source is incorrect or inaccessible. The image has been removed from the dissertation as it was not necessary to include.
17. Page 7: Katsasis and Jabs 2014 is missing from the reference list. I assume Table 1.1 was developed by the candidate and not sourced from this reference? The list of major and minor clinical features of TCS was adapted from Katsanis and Jabs, 2004. The reference was included in the reference list under the Books section. See Page 114
18. In Table 1.1 under Minor, correct ‘Cleft palate and/or palate’ - missing word? The statement has been completed to cleft lip and/or palate. Page 5
19. Page 8: Connor and Ferguson-Smith, 1988 is missing from the references list. A more recent estimate of genetic diagnoses for TCS would be preferable. A more recent estimate of 60% TCS de novo variants has been added (Splendore *et al*., 2002). Page 5, Paragraph 1
20. Original reference for HGMD is Krawczak and Cooper, 1997. Need to give URL (add to electronic sources) and mention date accessed for number of mutations reported for

TCS. The reference has been corrected throughout the dissertation and HGMD URL is given in electronic sources reference list, page 115.

21. Page 9: Figure legend (Fig 1.5); indicate mutational hotspots are in purple. Mutational hotspots are indicated, Fig 1.5 is now Fig 1.1
22. Fig 1.6 adjust as necessary to have full image and legend on same page. The figure and legend are adjusted. Figure 1.6 is now Figure 1.2
23. Page 10: correct 'RNA polymerase enzyme IIll' used in two places, should be III. Check elsewhere. RNA polymerase III has been corrected throughout the dissertation.
24. Page 11: second sentence, repeated. The sentence has been removed. Page 9, paragraph 1
25. Page 12: Fig 1.7 is referred to in the text as "Figure 1.8". Figure 1.8 has been removed from the dissertation.
26. Page 13: Line 5, 'the most' is repeated. The repeated words were deleted.
27. Table 1.2: Re-order features by decreasing frequency. Table 1.2 has been re-ordered.
28. Morgan 1910 does not mention the DHODH gene OR wing anomalies and malformed posterior limbs, not an appropriate reference. The statement has been removed.
29. HGMD has been previously introduced and referenced, not required to repeat. See above comment regarding HGMD. Is the number of mutations (for MS and TCS above) reported on the database from HGMD Professional? The public database does not provide summary information for total number of mutations. The statement has been corrected, the number of mutations reported for TCS, NS and MS is reported on the HGMD professional database.
30. Page 16: Reference Halonen et al. 2006 appears to be incorrect, please check reference list: Halonen *et al*., 2006 is the correct reference and has been added to the reference list. Page 13, Paragraph 1
31. HALOKEN K., HUKKI J., ARTE S.., et al. 2006. Mandibulofacial dysostoses-craniofacial anomalies. Hum Mol Genet, 57, 78-Unable to find the publication either way. The reference is corrected to "Halonen K, Hukki J, Arte S, Hurmerinta K (2006) Craniofacial structures and dental development in three patients with Nager syndrome. J Craniofac Surg 17: 1180-1187". The reference is included in the reference list and can be accessed online at https://www.ncbi.nlm.nih.gov/pubmed/17119427

32. Correct this sentence and reference “NS is even rarer than MS, estimated to affect 3 in 1 000 000 individuals (Halonen et al., 2006)”. On pg. 12, MS is said to affect 1 in 1 000 000. The sentence has been corrected to “NS is estimated to affect 3 in 1 000 000 individuals”. Page 13, paragraph 1
33. Page 17: Table 1.3 - see comment for Table 1.2 above. Features are now ordered by decreasing frequency in Table 1.2 and Table 1.3
34. Page 18: "Although there is no curative treatment for TCS, MS and NS, there are however, management strategies to improve the patient's quality of life and ease the burden on their families.” Repetition - see page 4, par 3. The repetitive sentence has been removed. Page 15, para 1
35. Page 20: Par 2 “While there are at least 30 FDs reported in literature...” requires a reference. A reference has been added. Page 17, para 2
36. Page 21: Table 1.4, correct abbreviation ‘AR’. Table 1.4 has been corrected
37. Page 22: “The clinical diagnosis of these disorders is further complicated by disorders which can be considered under their differential diagnoses.” Avoid repetition. The repetitive statement has been removed. Page 19, para 1

Materials and Methods

38. Page 24: NHLS should be written in full at first use. NHLS is now written in full. Page 21, para 1
39. See previous comment on differential diagnoses for FDs — some info should be presented in the Introduction. A paragraph about differential diagnoses investigated in this has been added in the Introduction chapter. Page 15, para 2
40. NB. Include page numbers for all references to ‘Appendices’. Correct page numbers have been included in all text references to appendices.
41. Page 25: Wits should be written in full at first use (and included in abbreviations). The University of the Witwatersrand is now written in full at first and has been added in the abbreviation list. Page 21, para 1
42. “Written informed consent for genetic testing was obtained from all patients, and specific written informed consent for the publication of photos was” word missing at the end. The sentence has been completed with the word ‘obtained’ and now reads

"Written informed consent for genetic testing was obtained from all patients, and specific written informed consent for the publication of photos was obtained". Page 22, para 1

43. There are two deceased patients — how were the consent process and samples handled for them? Need to elaborate. These were patients who had attended at least one of the Division's genetic clinics before they passed on. In those clinics, blood samples are normally taken for DNA banking and/or other genetic investigations. Similarly, with the two deceased patients included in this study, their blood samples were taken before they passed on and DNA was extracted for other genetic investigations. Following screening of patients file by Medical Geneticist, the two patients were identified as suitable participants for this study and their parents or legal guardians were invited for an information session where they gave written informed consent for the use of their children's DNA for genetic testing in this study.
44. A database is mentioned in the second paragraph; please specify the platform, name/use and access. A Microsoft Excel database was used and this was only accessible to individuals directly involved in the study as it was stored in google drive. Page 22, para 2
45. Page 26: "a ratio lower than 1.8 indicates possible contaminants." Re-write, indicates the 'presence of possible contaminants. The statement has been rephrased, page 23, para 2
46. Page 27: "used one custom designed gene panel to sequence genes associated with each study." Re-write as "used a 'single' custom designed panel to sequence the genes associated with each of the sub-studies." The statement has been rephrased, Page 24, para 1
47. "The gene panel included coding regions, intron-exon boundaries, untranslated regions and ten flanking bases of 49 genes known to cause, or interact with causative genes of, the three studies." Re-write "The gene panel included coding regions, intron-exon boundaries, untranslated regions and 10 flanking bases of 49 genes known to cause, or interact with causative genes across the three sub-studies." The statement has been rephrased, Page 24, para 1

48. “NGS is sensitive to the concentration of the start-up DNA.” Replace the abbreviation at the start of the sentence and re-write "concentration of the starting material.” Rephrased and the sentence now reads “The concentration of the starting material is crucial to the optimal performance of an NGS”. Page 24, para 2
49. Correct “Higher Sensitivity (HS)’, should be “High Sensitivity’. Higher sensitivity is corrected to high sensitivity. Page 24, para 2
50. Page 28: “The protocol is lengthy to complete in a single day..." remove this from the text and the legend for Fig 2.1, it is enough to state that the protocol was performed over three days. The sentence has been removed. Page 25
51. Use “purification” rather than “clean-up” and “experiment” rather than “sequencing run”. Purification and experiment have been used throughout the dissertation.
52. NB. There are several grammatical errors and typos throughout Section 2.2.4. Please proof-read and edit. Some measurements are written in words not figures, lacking consistency. Section 2.2.4 has been proof-read and edited for grammatical and typos error. Measurements less than and equal to ten are written out in words and the rest in numerals consistently.
53. Page 31: Write “QC’ in full at first use. Quality control has been written in full at first. Page 28, para 1
54. Page 32: “...SAMTools and variants were then called using SNIPPET algorithm using the SureCall software on default parameters.” Avoid repetition. The sentence has been rephrased and now reads “Post alignment, a QC step which included removal of duplicate reads, indel local realignment and base quality score recalibration was performed using the SureCall software on default parameters”. Page 29, para 1
55. Page 33: line 3 says “see Appendix” - which one? Page number? The ACMG codes have been removed from the Appendix list as permission to reproduce their tables and figures was not obtained.
56. Page 34: Table 2.2, a better descriptive title would be “Primers designed for validation of putative disease-causing variants by Sanger sequencing”. Indicate (5' to 3’) orientation for the F/R primers. Table 2.2 has been renamed and the 5’ and 3’ orientation has been indicated for all primers. Page 31

57. “All primers were PCR optimised and their products were separated using Agarose gel electrophoresis. Figure 2.4 below shows electropherograms of the seven set of primers as visualised and imaged using a UV transilluminator.” NB. It is the PCR that is optimized NOT the primers. Also Fig 2.4 is a collection of gel images NOT electropherograms and | assume you mean ‘of the PCR products’ NOT the set of primers - we don't visualize primers. Include product sizes in Fig1. The sentences have been rephrased and the product size is included in the figure. See page 90, Fig 3.29

24. Use of ‘UV’ as an abbreviation not written out first use. Ultraviolet has been written out in full at first. Page 90, para 1

Results

58. Page 36: “Patients recruited were patients clinically diagnosed and/or suspected with..." Re-write “Patients recruited were those clinically diagnosed and/or suspected to be affected by...” The statement has been rephrased, Page 33, para 1
59. MFDM is mentioned in the abstract but nowhere else prior to it appearing in the Results section. Please include in the introduction under a section on differential diagnoses. A paragraph on differential diagnoses, including MFDM has been included in the Introduction section. Page 15, para 1
60. Table 3.1: FRASC in full, NA is listed in abbreviations but not used in the table. Check ALL tables, many similar issues. NA has been removed in all tables. FRASC is used as a unique code for the study. It has no full definition.
61. Page 38: “A total of 12 candidate genes were identified to include...” Re-write “A total of 12 candidate genes were identified for inclusion...” The statement has been rephrased. Page 35, para 1
62. The sentence “The exact mechanism by which mutations in SF3B4...” should start a new paragraph. The statement starts on a new paragraph. Page 35, para 1
63. The discussion of *SMAD* genes is disjointed by the above sentence — keep the SF3B4 gene info and mechanism together and the same for the SMADs. The text has been broken into four paragraphs. Page 35, para 1-4

64. Page 39: Fig 3.1, correct 'forth peak' and in subsequent figures. Check legend, FRASC28 is referred to twice. Centre point for FRASC28 appears to fall under 300bp. The legend has been corrected and the peak correctly sized. Page 36
65. Page 40: Fig 3.2, similar to FRASC61, centre point is under 300 bp. Can you comment on what this might mean for subsequent analyses? A comment on what this could mean for subsequent analysis has been added. Page 37
66. Page 41: "Seeing that all three runs worked..." is VERY informal language. Rephrased to "With the PhiX being successful in all three runs, the quality metrics of each run was then evaluated (Table 3.2)". Page 38, para 2
67. Page 42: Table 3.2, the use of numerical superscripts for your key is not ideal; suggest you substitute letters since you have squared measures. In the table, more informal language "is predicted to go well'. Numerical superscripts have been replaced with letter superscripts and the sentence has been rephrased to "The cluster density range in which optimal sequencing is predicted". Page 39, Table 3.2
68. ",..a Small portion of target regions in FRASC27, FRASC28 and FRASC54 had a Q-score of ≥ 20..." Should this be <20? The top of the page states high-quality data with a Q score of ≥30. There are other places in the dissertation with similar errors (check Q-score and read depth). The Q-score has been fixed, = Page 39, para 1.
69. NB. For ALL tables presenting candidate variants for each patient, please keep the order consistent across tables for the tools listed. Would be useful to include the applicable ACMG codes with the classification in each table. The order is consistent and the ACMG codes have been added to each variant throughout the dissertation.
70. A summary of reads and read depth is included for some of your patients (FRASC23, FRASC25 on pgs 48 and 50 respectively) but not all. I find this repetitive since info is already in Table 3.3 - you could add average read depth for each sample to the table to simplify and only mention it in the text where relevant e.g. FRASC27 where coverage was inadequate for some regions. Paragraphs in the results section repeating read depth and the number of reads generated have been removed except for FRASC27.
71. Page 60: Fig 3.11, Mutalyzer url should be provided and referenced in electronic resources, also IGV. The references for these should be added to Table 2.1. Thereafter, they can be named as tools without the need to reference each time. The mutalyzer tool and its URL has been added to the electronic reference list .Page 30, Table 2.1.

72. Page 69: The ACMG classification in the text doesn't match the table. Corrected, the variant is classified as a VUS. Page 45, Table 3.6.
73. Page 70: FRASC52, text states two siblings affected but not shown in pedigree? Corrected, a single sibling presented with clinical features suggestive of TCS. Page 68, para 2
74. Page 75: correct legend for Fig 3.17. The legend has been corrected, Page 66, Figure 3.16
75. Page 77: The *POLR1D* variant named in the text does not match the table. The variant has been corrected.
76. Page 80: Table 3.17, CHD7 variant does not have a classification. Corrected. The *CHD7* is classified as likely benign. Page 80, Table 3.17
77. Page 81: Multiple words missing in the text on this page in particular.
    - "The variant lies in a non-conserved and is predicted..."
    - "... the variant has been observed in 18 and 2 apparently individuals..."
    - "... predicts that the variant's transcript a target of NMD."

Page 80 has been proof-read and all texts are completed

78. There are a few other places in the text presenting the results for different patients. Please proof-read and correct. The Results section was proof-read and corrected.
79. Page 86: stated range of figures is wrong. A correct range of figures is given. Page 87, para 1
80. Page 87: Table 3.19, add ACMG classifications in a final column. FRASCS59 is listed as 'African' vs the others 'Black African'? The ACMG classification is added to the table and all African patients are referred to as African throughout the dissertation.
81. Page 88: All figures should be bigger and the putative disease-causing variants shown more clearly, e.g. using a different colour font or arrows, etc. Figures were enlarged and an arrow used to show putative disease-causing variants. Page 89-90, Fig 3.26-3.28
82. NB. Check ALL figures for gene names in italics, also consistency in describing (a) and (b) in the figure. The labelling of a and b figures have been corrected and used consistently throughout the dissertation.

Discussion

83. Page 89: “Internationally, a clinical diagnosis of TCS, NS or MS is confirmed...” needs a reference. A reference has been added (Katsanis and Jabs, 2004). Page 91, para 1
84. Page 90: The clinical phenotype and diagnostic yield of TCS and CHARGE syndrome are discussed in this section. Furthermore, a speculation on each variant’s potential mechanism of action is also presented. The discussion is restricted to TCS and CHARGE syndrome...”Please re-write. The statement has been rephrased. Page 92, para 1
85. Would be good to somewhere remind the reader of the inheritance patterns for each of the disorders discussed. Inheritance pattern of TCS and CHARGE syndrome has been added in the beginning of their respective discussion section.
86. Page 91: Break up the wall of text into two or more paragraphs to make it easier to read. The text has been broken into three paragraphs. Page 102, para 2-4
87. Reiterate and comment on percent pick-up of TCOF and POLR7 gene putative mutations in this study compared to literature. A comment on the reported pick-up rate of the *TCOF1* and *POLR1D* has been added. Page 93, para 2
88. Third to last line, “which is located in one of the TCS mutation hot spots, exon 24.” Of which gene? The statement has been completed, Exon 24 of the *TCOF1* gene has been added. Page 94, para 1
89. NB. For each variant discussed, include a sentence at the beginning of the respective section stating its classification. The reader is otherwise forced to return to the respective tables in the Results section to look up the classification. A sentence stating the variant’s classification at the beginning of the respective section has been added.
90. Use appropriate referrals to your tables and figures in the Results section to back up the discussion. The discussion section now refers to appropriate Tables and Figures.
91. You mention skewing of results due to small sample size, can you comment on the sample size in the comparative studies. A Comment on the sample size in comparative studies is added. Page 92, para 3
92. Page 97: check use of italics for CHD7 protein. Italics were removed, Page 98, para 2

93. Page 99: “Also, testing of other family members, looking for the presence/absence of the putative familial disease-causing mutation, was also not performed due to time constraints.” The sentence has been rephrased and now reads “owing to time constraints, family studies could not be performed’’. Page 102, para 1
94. Page 100: see comment above page 91. The text has been broken into two paragraphs. Page 101, para 1-2
95. Page 102: Recommend careful editing of the conclusion section for grammar and repetition. The conclusion section has been edited.

References:

96. Please make a note of the referencing style used and the reference management tool at the beginning of the dissertation. The candidate used Harvard referencing style incorporated by the Reworks reference managing tool. Page Xi
97. There are numerous inconsistencies and errors in the references list, which should be corrected. Care must be taken to crosscheck the references list auto-generated by any tool. Formatting must be identical, italics for journal names and capital letters for all names e.g. Bouazoune et al, McCathy, etc. There are many instances of references listed in duplicate (and even triplicate) e.g. Chemke et al, Dixon et al, Franceschett et al, Splendore et al, etc. On page 113, reference for Wright et al starts at the end of Wieczorek. All in-text and references in the reference list were crosschecked and references in duplicate were removed.
98. The reference list should be in chronological order for publications with the same first author e.g. Bronner-Fraser, etc. Note that citations within the text should consistently be alphabetical if that is the chosen format. Currently they are sometimes chronological and sometimes alphabetical. The reference list is in chronological order with the same author and citations within the text have been corrected listed in alphabetical order throughout the references,